

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:36:26 ; Search time 4832.99 seconds
(without alignments)
11760.106 Million cell updates/sec

Title: US-09-882-986-1
Perfect score: 2716
Sequence: 1 aaagagacgaactgaagaaa.....ctttaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	2716	100.0	2716	6	AX341991	Sequence
2	2716	100.0	2716	8	AF272891	Arabidops
3	2700	99.4	2834	8	AF268377	Arabidops
4	2697.4	99.3	2735	8	AY064960	Arabidops
5	990.2	36.5	82403	8	AC007063	Arabidops
6	978.2	36.0	3009	8	AY017415	Oryza sat
7	69.2	2.5	65076	8	NCB13020	Neurospor
8	65	2.4	3228	3	DME238970	Drosophil
9	61.8	2.3	2975	3	AY051796	Drosophil
10	57.2	2.1	3066	9	HOMNCPB	Drosophil
11	56.8	2.1	14831	2	AC013864	Drosophil
12	56.8	2.1	160622	3	AC023715	Drosophil
13	56.8	2.1	167089	3	AC023747	Drosophil
14	56.8	2.1	311464	3	AE003431	Drosophil
15	56.2	2.1	163029	9	AC022016	Homo sapi
16	55.6	2.0	2541	9	BC001450	Homo sapi
17	55.6	2.0	2828	9	HSCBP80	H.sapiens C
18	53.2	2.0	28440	8	SPAC6G10	S.pombe chr
19	50.8	1.9	162770	3	AC008309	Drosophil
20	50.8	1.9	167475	3	AC016132	Drosophil
c 21	50.8	1.9	220936	2	AC020256	Drosophil
22	50.8	1.9	240910	3	AE003734	Drosophil
c 23	48.4	1.8	78396	9	AL590486	Human DNA
24	46	1.7	125020	9	AF429315	Homo sapi
c 25	46	1.7	293431	2	PFM113P4	Plasmodiu
26	45.6	1.7	372	3	AF162337	Mytilus g
c 27	45.4	1.7	603	8	AF464753	Pholidota
c 28	45.4	1.7	6242	6	AX347050	Sequence
29	45.4	1.7	74371	9	AC005369	Homo sapi
c 30	45.2	1.7	7218	6	I66494	Sequence 14
31	45	1.7	1386	6	AR075041	Sequence
32	45	1.7	1386	6	ARI41859	Sequence
33	45	1.7	2175	10	BC007144	Mus muscu
c 34	45	1.7	3780	3	AF238312	Dictyoste
35	44.8	1.6	125021	8	AC007504	Arabidops
36	44.6	1.6	9330	3	AF466305	Dictyoste
37	44.4	1.6	2062	3	AF035482	Trichoplu
c 38	44.4	1.6	6552	6	AX251432	Sequence
39	44.2	1.6	1761	10	BC022641	Mus muscu
c 40	44.2	1.6	6063	6	AX323780	Sequence
c 41	44.2	1.6	99053	8	AC005292	Genomic s
c 42	44.2	1.6	161619	9	AC046174	Homo sapi
c 43	44.2	1.6	167106	2	AC090102	Homo sapi
c 44	44.2	1.6	293431	2	PFM113P4	Plasmodiu
c 45	44.2	1.6	349980	6	AX344572	Sequence

ALIGNMENTS

RESULT 1
AX341991
LOCUS AX341991 2716 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1 from Patent WO0196585.
ACCESSION AX341991
VERSION AX341991.1 GI:18137972
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Schroeder, J., Hugouvieux, V. and Kwak, J. M.
TITLE Modulation of abscisic acid signal transduction in plants
JOURNAL Patent: WO 0196585-A 1 20-DEC-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (US)
FEATURES
Location/Qualifiers
1..2716
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/note="abscisic acid (ABA) hypersensitive (ABH1) cDNA"
37..2583

CDS

/note="ABH1 protein"
/codon_start=1
/protein_id="CAD20687.1"
/db_xref="gi:18137973"
/translation="MSNWKTLRLRIGEGPEXGTSDDYKDHLETCFVIRRETERSGD
OVLPELLOCAEOLPHKIPLYGLTGLIGLLENEDFVQKLVESVHANPQVALDSGNCNSI
RILLPMTSELSKVIOPASLIWVETLSSAATVDEEGNPSWQPOQADFYVICILS
SLPWSGEAEQVPOEIERVLVQIAYLSIRKNSSTGLFFHNGEFESLSAEKDFVE
DLIDRQSLASGWLKESRPHLSFEAQLVAGKHELPIKCMQEPSRPHSRAYS
GKQKHDAFRPQIRLRNIFPANKMEDVQPIDRFVVEYLVDLVLYLNGCRKECASY
MANLPVTRVEYLMAETLFSOILLPOPEKLYTLVIMDLCKALPGAPFAPVAVAGV
RALFKISLDLWESRTRILWFSHLSNFOFTWPMEEWAFVLDLKPWAKRPFVQBEIL
QREVLRSYDKIKQSIENATALELLPPKAGPNFMYSLSEKKEEQQLSALSRLV
KEKQARDMWIEETIYPVHGFEVTLIVQVTLIDIGSKSFTHLVTLVLERYGVFQFSK
LCPDNDKMLVSTYKNNVQMTAVADRMGYRLVSNQAIWVWFSNPNVDQFH
VSDQPEILGNALNTYNIISDLRKDISNITKNVLVAEKASARVLEAEASKLSLV
EGEVLGNPNKMKRLKSTVEKTEGAELSRESLEAKALLNARLAKSETVELLLLFQS
FLGVLEKRLPDPKTVKPSVODLKSIGNEDDKPSMDVDSENGPNPKSCEVGERQWCLS
TLGYLTATFRQVASEIWPHEKLESEVFSGEDVHPLFLQAISSALQJPUH"
BASE COUNT 778 a 532 c 628 g 778 t
ORIGIN

Query Match 100.0%; Score 2716; DB 6; Length 2716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 aaagagacgaactgaagaaacccctcgagaaagatgagcaatggaaactcttctc 60
DB 1 AAGAGACGACACTGAGAAAACCTCTCGGAAGAAGATGACCAATGGAAACTCTTCTC 60
QY 61 cttcgactcggggaagacgtgagtagcggcactctccgactacaaagaccatc 120
DB 61 CTTGCGATCGGCGAAAAGGACCTGAGTAGCGGACATCTCTCCGACTACAAAGACCACATC 120
QY 121 gagactgttctggtcattctagagaaatcgagcgttctctggagatcaagtgttgcct 180
DB 121 GAGACTGTGTTTCGGTGTCAATTCGTAGAGAAATCGAGCGTCTCGAGATCAAGTTTGCCT 180
QY 181 ttctactacaatgtgtaaacatgtcctcataagattcctctgtatggacttgatt 240
DB 181 TTTCTACTACAATGTCTGAACAAATGGCCCTATAAGATTCCCTTTGATGGGACTTTGATT 240
QY 241 ggtttgtgaacttgagaaatgaagattttgtccagaagtagtagaagatgtccacgt 300
DB 241 GGTGTTGTGAACTTGAGAAATGAAGATTGTCCAGAGCTAGTAGAAGTGTCCAGGCT 300
QY 301 aatttccaggtcgttagattctggcaactgcaacagtagtaccgtatattgcttgccttt 360
DB 301 AATTTCCAGGTCGCTTTAGATTCTGGCACTGCAACAGTATCCGTATATTGCTTCGCTTT 360
QY 361 atgacttccctgtgtgcagtaaggttattcaacctgcttcttgattgtcgtcttcgaa 420
DB 361 ATGACTTCCTGTGTGTGCAAGTAGGTATTCAACCTGCTCTCTTTGATTGTGCTTCGAA 420
QY 421 acattgctatcatctgctgccactactgtgatgaagagaaagaaatccatcatggcag 480
DB 421 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAGAGAAAGGAATCCATCATGGCAG 480
QY 481 ccacaagctgacttttactgtatcatcttgcagcctcccgctggggagagatcagaa 540
DB 481 CCACAAGCTGACTTTTTACGTTATATGATCTTGTCCAGCCTCCCGTGGGGAGGATCAGAA 540
QY 541 ctgcgtgagcaagtctctgatgagattgaagagtgattagttgggatacaagcttattg 600
DB 541 CTGCGCTGAGCAAGTCTCTGATGAGATTGAAAGAGTGTGTTAGTTGGGATACAAAGCTTATTG 600
QY 601 agcatccgaaagaattcttccacctctgggttaaacctttttcacacgagagaatttga 660
DB 601 AGCATCCGAAAGAATTTCTCCACCTCTGGGTAAACTTTTTTTCACACGGAGAAATTGAA 660
QY 661 agcagccttgcagagaagatttcgtggaggatctattggatgaattcagctctggctt 720
DB 661 CTTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACTGGAACAACT 1800

DB 661 AGCAGCCTTGACGAGAAGGATTTGCTGGAGGATCTATTGGATCGAATTCATGCTCTCTGGCT 720
QY 721 tccaatgatggaaaacttgaagcgttacctagcctcatctctcgtttgaaagctcaactc 780
DB 721 TCCATGGATGGAAACTTGAAGGCTACCTAGGCTCATCTCTGTTTGAAGCTCAACTC 780
QY 781 gttgctggaaagtttctagagctacgttccattaaatgtatggaaacaccagagtcacact 840
DB 781 GTTGCTGAAAGTTTCATGAGCTACGTCCCATTAATGTATGGAACAACCCAGTCCACCT 840
QY 841 tctgatcattcggagggcatacacagtggcaagcaaaagcatgatgatgacagagatatccc 900
DB 841 TCTGATCATTCGAGGGGCATACAGTGGCAAGCAAAAGCATGATGCTTTCACGAGATATCCC 900
QY 901 cagagaattcgtaggttgaataatttccagcaataaaaaaggagatgatacaaccaatt 960
DB 901 CAGAGAATTCGTAGGTGATATATTTCCAGCTAATAAATGGAGGATGTACACCAATT 960
QY 961 gatcgttttctgctggaggagattttgctgagatgtgctctctatttggaaatggatgctgg 1020
DB 961 GATCGTTCCTGCTGGAGGAGTATTGCTGGATGCTGCTTCTATTGTAATGGATGTCGG 1020
QY 1021 aaggagtgtgcacctacatggctaatacttccctgttatacatttcggtacgagatcttctatg 1080
DB 1021 AAGGAGTGTGCATCTCATATGGCTAATCTTCTGTTACATTTCCGTACGAGTATCTTATG 1080
QY 1081 gcagagacactattttctcagactactgctaccctccagccacattcaagactcttcat 1140
DB 1081 GCAGAGACACTATTTCTCAGATACTGCTGCTACCCAGCCACCATTTCAGACTCTTTAT 1140
QY 1141 tatcacactcgtgattatggatcttttgaagcctcttcgggtgctcttctcctgctgtgtt 1200
DB 1141 TATACACTCGTGATTATGGATCTTTGTAGGCTCTTCCGGGTGCTTTCCTGCTGTGTT 1200
QY 1201 gctggcgctgtcgtgcaactatttgagaaaaataccgacttagacatcggaaatccagagacg 1260
DB 1201 GCTGGCGCTCTTCTGTCACATATTTAGAAAAATATCCGACTTAGACATGGAAATCCAGGACG 1260
QY 1261 cgtctatcctcctggttttctcaccacttatccaaacttccaaattcctcgtcgtgggaa 1320
DB 1261 CGTCTATCTCTGTTTTCTCACCACTTATCCAACTTCCAAATTCATCTGCGCGGTGGAA 1320
QY 1321 gagtggcctttgttggatcttcccaagtgggcccctcaagcgtgtatttctgttcagag 1380
DB 1321 GAGTGGCTTTTGTGTGGATCTTCCCAAGTGGGCCCTTAAGCGTGATTTTGTTCAGGAG 1380
QY 1381 atttgcgaagagaagtacgctgtcttactgggataaaaaataagcagagacatcgaagt 1440
DB 1381 ATTTTGCAAAAGAGAAGTACGCTTTGCTTACTGGGATAAAAATTAAGCAGAGCATTTGAGAA 1440
QY 1441 gcgactgccctagaagaatttactctccaaaagctggtccgaatttttatgtattccttg 1500
DB 1441 GCAGCTGCCCTAGAAGAAATTTACTTCTCCAAAAGCTGGTCCGAAATTTATGTATCTCTTG 1500
QY 1501 gaagaagttaaaagagaaaacagagaagacagcaattgtcagccgaattgagcaggaagtc 1560
DB 1501 GAAGAAGTTAAAGAGAAAACAGAGAAACAGCAATTTGTCCAGCGAATTTGAGCAGGAAGTCC 1560
QY 1561 aaggaaaaacaaccgcacgtgacatgatagtgtggattgaaagaacgatatataccagtt 1620
DB 1561 AAGGAAAAACAACCCGACGTCGATGATGATGTTGAGTGGATTGAAAGAAGCATATATCCAGTT 1620
QY 1621 catggttttaagttactcttacaagtgtgtacagaccttacttgacatcggtatcaaaa 1680
DB 1621 CATGGTTTTGAAGTTACTCTTACAANTAGTGTACAGACCTTACTTGTACATCGGATCAAAA 1680
QY 1681 agtttcaactatttggctcactgctgagcgaatagtgcccaagtatttttcaaaagcttctgt 1740
DB 1681 AGTTTCACTCATTTGCTCACTGCTCTGAGGCGATATGCCAAGTATTTTCAAGCTTTGT 1740
QY 1741 cctgataacgataagcaggtgtatgtatctatctcaagtgtgatactactggaatacaaat 1800
DB 1741 CCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACTGGAACAACT 1800

QY 301 aattccaggctgctttagattctggaactgcaactgcaactgatactcgttatattgcttcgcttt 360
|||||
Db 301 AATTTCAGGTCGCTTTAGATTCTGGCAACTGCAACAGTATCCGTATATTGCTTCGCTTT 360
QY 361 atgacttccctgctgtgtagagagttattcaactgcttcttggatgtgctgttcgaa 420
|||||
Db 361 ATGACTTCCCTGTTGTGACAGTAAAGGTTATTCAACCTGCTTCTTTGATTGTCGCTTCGAA 420
QY 421 acattgctatcatctgcccactactgtgtagaagagaagaaatccatcatcattgagcag 480
|||||
Db 421 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAAGAGAAAGAAATCCATCATGCGCAG 480
QY 481 ccacaagctgaactttacgtttatgcatcttgcagctccctggtgggagagatcagaa 540
|||||
Db 481 CCACAAGCTGACTTTTACGTTTATGCATCTGTGCCAGCCTCCGCTGGGGAGGATCAGAA 540
QY 541 ctgcctgagcaagtctcctgatgagattgaaagtgtagttgggatacaagcttattg 600
|||||
Db 541 CTCGCTGAGCAAGTTCTGATGAGATTGAAAGAGTGTAGTTGGGATACAAAGCTTATTG 600
QY 601 agcatccgaagaaatcttccactctggttaaacctttttcacacggagaaatttgaa 660
|||||
Db 601 AGCATCCGAAGAATTTCTCCACTCTGGGTTAAACTTTTTCACAAACGGAGAAATTTGAA 660
QY 661 agcagcttgcagagaagattctgtggagatctattggatcgaaattcagctctctgct 720
|||||
Db 661 AGCAGCCTTGCAGAGAAGATTTCGTGGAGGATCTATTGGATCGAATTCAGTCTCTGGCT 720
QY 721 tccaatggattggaactggaacgctaccttaggcctcctcctcctcctcctcctcctcctc 780
|||||
Db 721 TCCAAATGGATTGGAATCTGAAACGCTACCTAGGCTTATGATGATGATGATGATGATGATG 780
QY 781 gttgctggaagtttctcagctacgtcccatataatgtagaacaacagagctccact 840
|||||
Db 781 GTTCTGGAAAGTTTATGAGCTACGTCCTCCATTAATGATGATGATGATGATGATGATGATG 840
QY 841 tctgatacttcaggggcatcacagtggcaagcaaaagcatgagcattgacgagatcccc 900
|||||
Db 841 TCTGATCATTCGAGGGCATACAGTGCAAGCAAAAGCATGATGCATTCGACGAGATATCCC 900
QY 901 cagagaattcgtggttggaataatttccagctaaataaataagagagatgtacaaccaatt 960
|||||
Db 901 CAGAGAAATCGTAGGTGGAATATATTTCCAGCTAATAAATGAGGATGTACAACCAATT 960
QY 961 gatcgttttctgtagaagatttctgtagatgtcctcttctatttgaatgagatgcgg 1020
|||||
Db 961 GATCGTTTTCTGCTGGAGGAGTATTTGCTGGATGTGCTCTTCTATTATTTGAATGGATGTCGG 1020
QY 1021 aaggagtgtcatcctacatcgtgtaactcttccgtttacatttccgttacgagatctctatg 1080
|||||
Db 1021 AAGGAGTGTGATCCTACATGCTGCTAATCTTCTGTTACATTTCCGTTACGAGTATCTTTATG 1080
QY 1081 gcagagacactatttctcagatactgctgtcctacccagccacattcaagactctttat 1140
|||||
Db 1081 GCAGAGACACTATTTCTCAGATACTGCTGTACTACCCAGCCACCATTCAGAGACTCTTTAT 1140
QY 1141 tatcaactcgtgattatgacttttgaagctcttccgggtgccttctcctgctgtt 1200
|||||
Db 1141 TATCACTCTGTGATTAATGAGATCTTTGTAAGGCTCTTCCGGGTGCTTCTCTGCTGTGTT 1200
QY 1201 gctggcgctgttcgtgactatttgagaaataccgacttagacatggaaatccaggagc 1260
|||||
Db 1201 GCTGGCGCTTCTGCTGACTATTGAGAAATATCCGACTTAGACATGGAAATCCAGAGCG 1260
QY 1261 cgtctatcctcgtggttttctcaccactatccaacttccaattcactcgtggcggtggaa 1320
|||||
Db 1261 CGTCTTATCCTCTGTTTCTCACCACCTTATCCAACCTTCCAATTCATCTGCGCGTGGAA 1320
QY 1321 gactggccttttctgtagatcttcccaagtggccctcaagcgtgatttcttccagag 1380
|||||
Db 1321 GAGTGGCCTTTGTGTGGATCTTCCCAAGTGGGCCCTTAAGCGGTGATTGTGTTTCAGGAG 1380

QY 1381 atttgcagaagaagtacgcttcttacttggtgataaaataagcagcagcattgagaat 1440
|||||
Db 1381 ATTTTGCAAAAGAGAAAGTACGCTTGTCTACTGGGATAAAATTAAGCAGAGCATTCAGAAT 1440
QY 1441 gcgactcccttagaagaattacttctcccaaaaagctggctcagaatttttatgtattccttg 1500
|||||
Db 1441 GCGACTCCCTTAGAAGAAATTTACTTCTCCAAAAGCTGGTCCGAAATTTTATGTATTTCCTTG 1500
QY 1501 gaagaaggttaaaagaaaaacagaaagacagcaattgctcagccgaattgagcaggaagtc 1560
|||||
Db 1501 GAAGAAGGTAAAGAGAAAAACAGAAACAGCAATTTGTACGCCGAATTTGAGCAGGAAGTCT 1560
QY 1561 aaggaaaaaaacaacccgcacgtgacatgtagtggattggaagaacatatatccagtt 1620
|||||
Db 1561 AAGSAAAAACAAACCGCACGTCACATGATAGTGTGGATTGAAGAAACGATATATCCAGTT 1620
QY 1621 catggtttgaagtacttcttaataatagttgtacagaccttacttgacatcggatcaaaa 1680
|||||
Db 1621 CATGGTTTTGAAGTTACTTCTTACAAATAGTTGTACAGACCTTTACTTGACATCGGATCAAAA 1680
QY 1681 agtttcaactcatttgcactgtcctgagcgatatggccaagtattttcaaaagctttgt 1740
|||||
Db 1681 AGTTTCACTCATTTGGTCTACTGCTCTGGAGCGATATGGCCAAGTATTTTCAAAGCTTTGT 1740
QY 1741 cctgataacgataagcagggtgacttattctcaagtgtgacatactgacatactggaaaaacaat 1800
|||||
Db 1741 CCTGATAACGATAAGCAGGTGATGCTATTATCTCAAGTGAGTACATACCTGGAACAAACAT 1800
QY 1801 gtacaaatgacggcggtggcaattgatatgagatgaggggtatagactagatcattcaatcag 1860
|||||
Db 1801 GTACAAATGACGGCGGTGGCAATTTGATAGGATGATGGGTTATAGACTAGTATCTAATCAG 1860
QY 1861 gcaattgttagatgggtgttctctccagaaaaatgttgateagtttcaagtctgctgatacag 1920
|||||
Db 1861 GCATTTGTTAGTGGGTTCTCTCCAGAAAATGTTGATCAAGTTCATGTCTGCTGATCAG 1920
QY 1921 ccatgggagatacttggcaatgctcttaacaagacttataaccgctatctcctgatttgagg 1980
|||||
Db 1921 CCATGGGAGATACTTGGCAATGCTTTAAACAAGACTTATAACCGTATCTCTGATTGTAGG 1980
QY 1981 aaagatatatacaacattacgaaaaaatgttttggctgagaaaaagcttccagcaatcaca 2040
|||||
Db 1981 AAAGATATATCAACATTTACGAAAAATGTTTGTGTTGCTGAGAAAGCTTCAGCCCAATGCA 2040
QY 2041 cgagttagatttgagcgtgctgagagcaaaactttccctagtggaagtggaacccgtctct 2100
|||||
Db 2041 CGAGTAGATTGGAGGCTGCTGAGAGCAAACTTTCCCTAGTGGAAAGGTGAACCCGTTCTT 2100
QY 2101 ggtgagaaatccagcgaagatgaagcgttttaaaatcaacagtggaagacacaggggaagcg 2160
|||||
Db 2101 GGTGAGAAATCCAGGAAAGATGAAGCGTTTAAATCAACAGTGGAGAACACAGGGAAGCG 2160
QY 2161 gagttactctcctcgggagtccttagagcaaaaagagcctcttcttaacagagctctctct 2220
|||||
Db 2161 GAGTTATCTCTTCGGGAGTCCCTAGAGCAAAAAGAGGCTCTTCTTAACAGAGCTCTCTCT 2220
QY 2221 gaaacccaggttttactgctcttctgcttccaaagtttcttagtggtgactgaagaaacgg 2280
|||||
Db 2221 GAGACCCAGAGTTTACTGCTCTTCTGCTGCTCCAAAGTTCTTAGGTGTACTGAAGGAACGG 2280
QY 2281 ctcccagatccaactaaagtagatcagtgagagatctaaatctatagggtgctgagaat 2340
|||||
Db 2281 CTCCCAGATCCAACATAAGTGAGATCAGTGCAGGATCTAAAAATCTATATAGGTGCTCAAGAT 2340
QY 2341 gacaagccatctgagatggagcgtgagacgagagaatggaaaaaccccaaaagagttgcgaa 2400
|||||
Db 2341 GACAAGCCATCTCGATTCGATTCGACGTGACAGCGAGAAATGGAAACCCCAAAAGAGTTGCCAA 2400
QY 2401 gtcggtgagagaaacagtggtgcttatacaacttggctatctcacggcattttacaagg 2460
|||||
Db 2401 GTCGGTGAGAGAAACAGTGGTGTCTTATCAACACTTGGCTATCTCACCGCATTTACAAGG 2460
QY 2461 caatgtagcagcagagatatggcctcacatggagaagttggagtagcagaagtgcttcggt 2520


```
|||||
Db 2461 CAATATCGCAGGAGATATGGCTCACATGGAGAAGTTGGAGTCAGAAAGTTCTTCGGGT 2520
QY 2521 gaagatgtgcatctctcttctccaaagcattctctctgcaacttccattccattacat 2580
Db 2521 GAAGATGTGCATCCTCTCTTCTCCAAAGCCATATCTCTGCACATCAATTCCTCCATATACAT 2580
QY 2581 taatctctctcttcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 2640
Db 2581 TAATCTCTCTCTTCAATCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCT 2640
QY 2641 gattctgacatcaagtatttaggaattgaaagagtcacaaagagtcacaaagagtcacaaagagtc 2700
Db 2641 GATTCTGACATCAAGTATTATTAGGAATTTGAAAGAGTCACAAAGAGTCATTAAGACTTT 2700
QY 2701 aaaaaaataaaaaa 2716
Db 2701 AAAAAAAAAAAAAA 2716

RESULT 3
AF268377
LOCUS Arabidopsis thaliana nuclear cap-binding protein CBP80 mRNA, PLN 14-JUN-2000
DEFINITION complete cds.
ACCESSION AF268377
VERSION AF268377.1 GI:8515770
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 2834)
Kmieciak, M. and Jarmolowski, A.
A nuclear cap-binding protein complex from Arabidopsis thaliana.
Molecular characterization of Arabidopsis thaliana CBP80
Unpublished
2 (bases 1 to 2834)
Kmieciak, M. and Jarmolowski, A.
Direct Submission
Submitted (16-MAY-2000) Gene Expression, A. Mickiewicz University,
Institute of Molecular Biology and Biotechnology, Miedzichodzka 5,
Poznan 60-371, Poland
FEATURES
Location/Qualifiers
1..2834
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/dev_stage="ten day old"
/note="whole plant"
170..2716
/note="similar to Homo sapiens CBP80"
/codon_start=1
/product="nuclear cap-binding protein CBP80"
/protein_id="AA076167.1"
/db_xref="GI:8515771"
/translation="MSNWKTLRLRIGRKEPYGTSSDYKDIETCFGVRIRREISGD
QVLPFLQCAQOLPHKIPLYGTLLGLLENEDFVKLVESVHANFOVALDSGNCNSI
RILLRFMTSLDKVIOPASLIVVFETLLSSAATVDEKGNPSQPOADVFICILS
SLPWGGSLEAQVDEIERVLVGLQIAYLSIRNSSTGLNFHNGEFPSSLAERKDFVE
DLLRIQSLASNGWKLESVRPHLSFEAQLVAGKFHELRPKCMQPPSPDSHRAYS
GKQKHALDVFYRQIRLNIPFANKMEDVQPIDFRVVEEYLDLVFLYNGRKECASV
MANLPVTIRYFVLAETFLSQILLPQPFKTYLYTLVIMDLKALPQAFPAVAVAGV
RALFEKISDLDMESRRLILWFSHLSNFORIPWPEWAFVLDLPKWPAPKRVFVEIIL
QREVLRSYWDKIKOSIENATALEELLPPKAGPNFMYLSEKEKEKEQQLSAELSRKV
KEKQATRMIVVIEETIYPVHGFEVTLIVVOTLIDIGSKSFTHLVTVLERYQVFSK
LCPNDKQVMLLSQVSTYKVNQMVTAVIDRMGMVRLVSNQAIVRVWFSPENVDHF
VSDQPVLEGNALNKYRISDLRKDISNITKNVLVAEKASANARVLEAEKSLSV
EGEPVLGNPAMKRLKSTVEKTEGAELSLRLEAKALLNRLALSEVFLILLFQS
FLGVLKERLPDPTKVRSVQDLKSI GAEDDKPSAMDVDSENGNPKSCVEGEREQWCLSL
TLGLVITAFTRQYASEIWPMEKLESEVFGSDVHPLFLQAISSALQFPPLH"
```

```
BASE COUNT 823 a 553 c 650 g 808 t
ORIGIN
Query Match 99.4%; Score 2700; DB 8; Length 2834;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2700; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 aaagagacgaactgaagaaaaaacctctcgggaagatgagcaattggaaaaactctctc 60
Db 134 AAAGAGAGCAACTGAAGAAAAAACCTCTCGGAAGAAGATGAGCAATTTGGAAAACTCTTCTC 193
QY 61 ctccgcatcggcgaaaaaggacacctgagtagcgcaactctccgactacaaagacacatc 120
Db 194 CTTTCGCATCGCGGAAAAAGGACCTCGTAGCGGCACTTCTCTCCGACTACAAAGAGACCACATC 253
QY 121 gagactgtttcgtggttcattcgttagagaaatcgagcgttctcgagatcaagttttgctt 180
Db 254 GAGACTTGTTGGTGTCTATTCGTAGAGAAATCGAGCGGTCTGGAGATCAAGTTTGGCT 313
QY 181 ttctactacaatgtcgtgaacaaattgcctcataagattcctcttgatggactttgatt 240
Db 314 TTTCTACTACAATGTGCTGAACAATTCCTCATAAAGATTCTCTTGTATGGGACATTTGATT 373
QY 241 ggtttgtgaacttggaagaaatggaattttgtccagaagctagtagaagaagtgtccacgct 300
Db 374 GGTTTGTTGAACCTGGGAGAAATGAAGATTGTCGAGAAAGCTAGTAGAAGTGTCCACGCT 433
QY 301 aatttccagctcgcttttagattcttgcaactgcaacagatccgtatatattcttcgcttt 360
Db 434 AATTTCCAGGTCGCTTTAGATTCTGGCAACTGCANCAAGTATCCGTATATTGCTTCGCTTT 493
QY 361 atgacttccctgtgtgacgtgaagttattcaacctgcttctttgattgctgctctcgaa 420
Db 494 ATGACTTCCCTGTTGTGCACATAAGTTATTCAACCTGCTTCTTTGATTGTCTCTCGAA 553
QY 421 acattgctatcatctgctgcactactgctggaagaagaaagaaatccatcatgagcag 480
Db 554 ACATTGCTATCATCTGCTGCCACTACTGTGGATGAAGAGAAAGGAAATCCCATCATGTCGAG 613
QY 481 ccacaagctgacttttacgttatatgcatcttccagcctcccggtgggagagtcagaa 540
Db 614 CCACAAGCTGACATTTACGTTATATGCACTCTTGTCCAGCCTCCCGTGGGAGGATCAGAA 673
QY 541 ctgcgtgagcaagtctcctgagatgagattgaaagagtgtagttgggatacaagcttatttg 600
Db 674 CTCGCTGAGCAAGTCTCTGATGAGATTGAAAGAGTGTAGTTGGGATACAAGCTTATTGT 733
QY 601 agcatccgaaaagaattctccacctctggttaaaactttttccacaacggagaatttgaa 660
Db 734 AGCATCCGAAAGAAATTCTTCCACCTCTTGGGTAAACCTTTTTCACAAACGGAGAAATTTGAA 793
QY 661 agcagccttcagagaagaggttcctgaggagatctattgagatcgaaatcagctctgct 720
Db 794 AGCAGCCTTCAGAGAAGGATTTCGTGGAGAGATCTATTGGATCGAATTCAGCTCTGCTGCT 853
QY 721 tccaatggatggaaacttgaacgctacacctcctcctcctcctcctcctcctcctcctcctc 780
Db 854 TCCAATGGATGGAACCTTGAAGCGGTACCTTAGGCTCATCTCTCGTTTGAAGCTCAACTC 913
QY 781 gttcgtgaaaagtcttcagatgactacgtcccatataatgtagaacaacogagtcacacct 840
Db 914 GTTGTCTGGAAGTGTTCATGAGCTACGTGCCCATTAATTAATGATGGAACAACCGAGTCCACCT 973
QY 841 tctgatacttcaggagcatcacagtgcgaacaaagcatgatgcattgagagatatccc 900
Db 974 TCTGATCATTTTCGAGGGGCATACAGTGGCAAGCAAAAGCATGATGATTCAGAGAGATATCCC 1033
QY 901 cagagaattcgtaggttgaaatataattccagactaaataaataagagagatgtacaacacct 960
Db 1034 CAGAGAATTCGTAGTGTGAATATATTTCAGACTAATAAATATGAGGAGATGTACACACCAATT 1093
QY 961 gatcgtttttgctgaggagagatatttgctggatgtgctctctctctattgaatggatgctgg 1020
```

Db 1094 GATCGTTTGTGCGGAGGATATTGCTGGATGCTCTTCTATTGAATGGATGTCGG 1153
Qy 1021 aaggagtgatcatctacatgctaaatctctctcttctacattctcgtacgaatattctatg 1080
Db 1154 AAGAGGTGTCATCCTACATGGCTAAATCTTCTGTATACATTCGGTAGAGTATCTTATG 1213
Qy 1081 gcagagacactattttctcagatactgctgtacccagccaccattcaagactctttat 1140
Db 1214 GCAGAGACACTATTCTCAGATACTGCTGTACCCAGCCACCATTCACAGACTCTTTAT 1273
Qy 1141 tatacactcgtgattatggatctttgttaaggctctcttcogggtccttccctcgtctgttt 1200
Db 1274 TATACACTCGTGATATTGATGATCTTTGAAGGCTCTTCCGGGTGCTTTCCCTGCTGTGTT 1333
Qy 1201 gctggcgtgctgctgacatttgagaaatatccgacttagacatggaatccagagacg 1260
Db 1334 GCTGGCGCTGTCTGTGACNATTTTGGAGAAATATCCGACTTAGACATGGAAATCCAGGAGC 1393
Qy 1261 cgtcttatcctcgtggtttctccaccacttatccaaacttcccaattcatctggtccgtggaa 1320
Db 1394 CGTCTTATCCCTCTGTTTCTCACCACCTTATCCAATTTCCAATTCATCTGGCGGTGGAA 1453
Qy 1321 gagtggccttttggttggatctctcccaagtgggcccctaaagcgtgtatttggctcagag 1380
Db 1454 GAGTGGGCTTTTGTGTTGGATCTTCCCAAGTGGGCCCTTAAGCGGTGATTTTGTTCAGGAG 1513
Qy 1381 atttgcagagagagacgtcttcttactcgtggataaaatgaagcaagcattgagat 1440
Db 1514 ATTTTGGAAAGAGAAGTACGCTGTCTTACTTGGGATAAAATTAAGCAGAGCATTTAGAAAT 1573
Qy 1441 gcgactgccttagaagaattacttctccaaagctgctcgaattttatgtattccttg 1500
Db 1574 GCGACTGCCCTAGAGAATTTACTTCCCAAAGCTGCTCGAATTTTATGTATCTCTTG 1633
Qy 1501 gaagaaggttaagaaacagagaaacagcaaatgtgtcagccgaattgagcaggaagtc 1560
Db 1634 GAAGAAGGTAAGAGAAACAGAGAAACAGCAATTTGTGAGCGCAATTTGAGCGGAAGTGC 1693
Qy 1561 aagaaaaaacacgcagcgtgacatgtagtggattggaagaaacatatatccagtt 1620
Db 1694 AAGAAAAAACACGCGACGTGACATGATGTGTGGATTGAGAAACCATATATCCAGTT 1753
Qy 1621 catggtttgaagtactcttacaatagttgtacagaccttactgacatcggatcgaataa 1680
Db 1754 CATGTTTGAAGTACTCTTACAATAGTTGTACAGACCTTACTTGACATCGGATCAAA 1813
Qy 1681 agttcactcatttggctcactgtcctggagcgatagtgcccaagtattttcaagctttgt 1740
Db 1814 AGTTTCACTCATTTGGTCACTGTCTCTGGAGCGATATGGCCAAAGTATTTTCAAAGCTTTGT 1873
Qy 1741 cctgataacgataagcaggtgatgcttatatactcaagtgaagtacatactggaatacaat 1800
Db 1874 CCTGATAACGATAAGCAGGTGATGCTATATCTCAAGTGAGTACATCTGGAATAACAAAT 1933
Qy 1801 gtacaaatgcgcggtggcgaattgatagtgatgggttatagactagttacttaacag 1860
Db 1934 GTACAAATGACGGCTGGCAATTTGATAGGATGATGGTTATAGACTAGTATCTAATCAG 1993
Qy 1861 gcaattgttagatgggtgttctctccagaaaaatgttgatcagtttcattcgtctgatcag 1920
Db 1994 GCAATTGTTAGATGGGTGTTCTCTCCAGAAAAATGTTGATCAGTTTCATGTGCTGATCAG 2053
Qy 1921 ccattgggagatacttgggcaatgctcttaacgaactataacccgtatctcgtattgag 1980
Db 2054 CCATGGGAGACTTGGCAATGCTCTTAACAGACTTATAACCGTATCTCTGATTGAGG 2113
Qy 1981 aaagatatatacaacattacaaaaatgtttgttgctgagaaagcttcagcccaatca 2040
Db 2114 AAAGATATATCAACATTACGAAAAATGTTTGGTGTGAGAAAGCTTCAGCCCAATGCA 2173
Qy 2041 cgagttagagttggagcgtcgtgagagcaaaccttccctcagtggaaggtgaaaccccttctt 2100

Db 2174 CGAGTAGAGTTGGAGGCTCTGAGAGCAAACTTTCCCTAGTGGAAAGGTGAACCCGTTCTT 2233
Qy 2101 ggtgaagatccagcgaagatgaagcgtttaaaatcaacagtggaagaagcaggggaagcg 2160
Db 2234 GGTGAGAAATCCAGCGAAGATGAAGCTTTAAATCAACAGTGGAGACACAGGGGAAGCG 2293
Qy 2161 gagttatctctcgggagtcctcctagaggcaaaagggtctcttcttaacagagactctctct 2220
Db 2294 GAGTTATCTCTCGGAGTCCCTAGAGGCAAAAGAGGCTCTTCTTAAACAGAGCTCTCTCT 2353
Qy 2221 gagaccgaggttttactcctctctgtgttccaaaatttcttaggtgtactgaagaaacg 2280
Db 2354 GAGACCGAGGTTTACTGCTCTTGTCTTCCAAAGTTCTTTAGGTGTACTGAAGGAACGG 2413
Qy 2281 ctccagatcccaactaaagttagatcagtcgaggtctcaaaatctatagtgctgaagat 2340
Db 2414 CTCCAGATCCAACTAAAGTGAGATCAGTGCAGGATCTAAATCTATATAGTGTCTGAAGAT 2473
Qy 2341 gacaagccatctgcgatggagcgtggacagcgagatggaaaccccaaaagaggtgtgcgaa 2400
Db 2474 GACAAGCCATCTGCGATGGACGTGGACAGGAGAAATGGAACCCCAAGAGAGTTGCGAA 2533
Qy 2401 gtcggtgagagagaacagtgtgtttatcaacacttggctctatctcacgcatttacaag 2460
Db 2534 GTCGCTGAGAGAGAACAGTGTGTCTTATCAACACTTGGCTATCTCACGCCATTTTACAGG 2593
Qy 2461 caatgcgagcgagatagtgccctcacatggagaagttgagtcagaagtgttctcgggt 2520
Db 2594 CAATATGCGAGCGAGATATGGCTCAGTGGAGAGTTGGAGTCAGAAGTGTCTCGGCT 2653
Qy 2521 gaagatgtgcattcctctcttctccaagcattcttctgcacttcaatccattccattacat 2580
Db 2654 GAAGATGTGCATCTCTCTTCTCCAAGCATATCTTCTGCACCTTCAATTCATTTCCATATCAT 2713
Qy 2581 taattctcctcttcaatctcaatcaaacctgtctcttcttcttcttcttcttcttcttct 2640
Db 2714 TAATCTCTCTCTTCAATCTCAATCAAACTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 2773
Qy 2641 gattctgacatcaagttattagaaattgaaagagtcataaaacaaagagtttcaacttt 2700
Db 2774 GATTCTGACATCAAGTTATTAGAAATTTGAAAGAGTCAAAAACAAAGAGTTTAAACTTT 2833

RESULT 4
AY064960 2735 bp mRNA linear PLN 10-DEC-2001
LOCUS Arabidopsis thaliana At2g13540/T10F5.8 mRNA, complete cds.
DEFINITION Arabidopsis thaliana
ACCESSION AY064960
VERSION AY064960.1 GI:17473508
KEYWORDS FLI CDNA.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 2735)
Cheuk R., Chen H., Kim C.J., Meyers, M.C., Banh, J., Bowser, L.,
Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,
Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K.,
Yanamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A.
and Ecker, J.R.
TITLE Arabidopsis cDNA clones
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2735)
Cheuk R., Chen H., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L.,
Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,
Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K.,

Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A.
and Ecker, J.R.

TITLE JOURNAL

Submitted (30-NOV-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEN (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H.,
Kim, C.J., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Dale, J.M.,
Goldsmith, A.B., Jiang, P., Jones, T., Karlin-Neumann, G., Lam, B.,
Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S.,
Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C.,
Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W.,
Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers

1. 2735

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="2"

/clone="RAFL09-32-L06(R21227)"

/note="ecotype: Columbia"

1. 70

71..2617

/note="putative cap-binding protein"

/codon_start=1

/product="At2g13540/T10F5.8"

/protein_id="AAL38378.1"

/db_xref="GI:17473509"

/translation="MSNKKVLLLRIGKEPPEYSSDYKHIECTFCGVRREIRBSGD
QVLPFLQCAOLPKIPLYGLTGLLLENEDFVQLVESVHANFQVADSGNCSI
RILLFRLSLCKVIQIPASLIVFETLLSSAATVDEEKNPSWQADPFYVICILS
SLPMGSELAEQVDEIRVLGVQAYLSIRKNSSTGLNFHNGEPFESSLAEDFVE
DLIDQISLASNGKLESPRPHLSFAQLVKGFLHRLPIKMEOPSPSDHSHRAYS
GKOKHDALTYPQIRIRNLINPANKMEDVOPIDRFVVEEYLLDLVFLNGRKCASV
MANLPVFRLEYLAETLFSQILLPQPPKLYLYLIMDLKALCPAFVAVAGV
RALFEKISLDMESTRLLIMFSLHLSNFQIWPWEWAFVLDLPKAPKRVFOEIL
OREVRLSYWDKIKOSIENATALELLPPKAGPNFMYSLIECKERTEQQLSAELSRKV
KEKQATDMIVWIEETIYPVHGFVETILTIVQTLIDIGSKSFTHLVLRYGQVFSK
LCPDNDKQVLLSOSTYKNNVQMTAVADRMGYRLVSNQATVRVVESPENVDOFH
MSDPWEILGNALNKTYNRLSDLEKDISNITKNVLYAEKASANARVELEAESKLSLV
EGEPVLGENAPKMKRLKSTVEKTEGEALSRESLEAEKALLNRALSTEVETLILLFOS
FLGVKLERLPDPTKVRSDQLKSGIAGEDDPSAMDVDSNGNPKKSCVEGEREQWCLS
TLGYLTAFTQYASEIWPHEKLESEVFSGEDVHPLFLQALSSALQPLH"

misc_difference 1943

/note="compared to genome"

/replace="g"

2618..2735

778 a 539 c 637 g 781 t

3'UTR

BASE COUNT

ORIGIN

Query Match 99.3%; Score 2697.4; DB 8; Length 2735;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2698; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 aaagagacgaactgaagaaacctctcggaagagatgagcaattggaactcttc 60

35 AAAGAGACGAACACTGAAGAAAACCTCTCGGAAGAAGATGAGCAATTTGAAAACCTTTCTC 94

61 ctccgcatcggaagagggacctgagtcaggcactctccgactacaaagaccacatc 120

95 CTTGCGCATCGCGAAAGAGGACCTGAGTACGCGACTTCTCCGACTACAAAGACCACATC 154

QY 121 gagactgttggtgtcattcgtagagaaatcgagcgttctcgagatcaagtttgcct 180
DB 155 GAGACTGTGTTGGTGTCAATTCGTAGAGAAATCGAGCGTCTCGAGATCAAGTTTGCCT 214
QY 181 ttctactacaatgtgtgaacaattgcctcataaagattccttctgtatggacttgcatt 240
DB 215 TTCTACTACAATGTCTGACAAATTCCTCATAGATTCCTTTGTATGGAGACTTTGATT 274
QY 241 ggtttgtgaacttgagaaatgaagattttgtccagaagctagtagaagtggtccacgct 300
DB 275 GGTTTGTTGAACCTTGGAGAATGAAGATTGTTGCCAAGCTAGTAGAAGCTGTCCACGCT 334
QY 301 aatttccaggtgcgttagatttctggcaactgcaacagtatccgtatatattgcgttc 360
DB 335 AATTTCCAGGTGCCTTTAGATTCTGGCAACTGCAACAGTATCCGTATATTTGCTTCGCTTT 394
QY 361 atgacttccctgtgtgcagtaagggttatcaacctgcttcttctgtatgctcgttcgaa 420
DB 395 ATGACTTCCCTGTTGTGCGAGTAGAGTTATTCAACCTGCTTCTTTGATTGTCGCTTCGAA 454
QY 421 acattgctatcctgtctccactactgtggtatgaagagaagaaatccatcatggtcag 480
DB 455 ACATTGCTATCATCTCTGCCACTACTGTGGATGAAGAGAAAGAAATCCATCATGCGCAG 514
QY 481 ccacaagctgaacttttaactgtatatgtatctgtccagcctccctggtgggagagatcagaa 540
DB 515 CCACAAGCTGACTTTTACGTTATATGCAATCTTGTCCAGCTCTCCGCTGGGAGGATCAGAA 574
QY 541 ctccgtgagcaagttcctctgagattgaaagtggttagtggtgggatacaagcttatttg 600
DB 575 CTCGCTGACCAAGTTCCTGATGAGATTGAAGAGTGTAGTTGGGATACAAAGCTTATTG 634
QY 601 agcatccgaagaattcttccacctctggtttaaactttttcacacacgagagaatttgaa 660
DB 635 AGCATCCGAAGAAGATCTTCCACCTCTGGGTAAACTTTTTCACAACGGAGAATTTGAA 694
QY 661 acagccttgagagaagattctgtggaggtatctattggatgagaattcagctcgtcggct 720
DB 695 AGCAGCCTTGCAGAGAAAGGATTTCTGTTGGAGGATCTATTGGATCGAATTCAGTCTCGGCT 754
QY 721 tccaatggtatgaacttgaaagcgtcacctagcctcctctctcgtttgaagctcaactc 780
DB 755 TCCATGATGAAACTTGAAGAGCTACCTAGGCCTCATCTCTCGTTTGAAGCTCAACTC 814
QY 781 gttgtgaaaagtcttcagagctacgtcccatataatgtatggaacacacgagctccacct 840
DB 815 GTTGCTGAAAAGTTTCATGAGCTAGCTCCCATTAATGTATGGAACAACCGAGTCCACCT 874
QY 841 tctgatactcgagggcatacagtggaagcaaaagcatgcatgcatgagagatcccc 900
DB 875 TCTGATCATTCAGGGGCATACAGTGGCAAGCAAAAGCATGATGCAATTGACGAGATATCCC 934
QY 901 cagagaattcgtggtgaataatttccagctaatcaaaatgagagatatacaaccaatt 960
DB 935 CAGAGAAATTCGTAGGTTGAATATATTTTCAGCTATAAATGGAGGATGTACAACCAATT 994
QY 961 gatcgttttctgctgagagagatttctggtggtgctctctatttgaatggatgcgg 1020
DB 995 GATCGCTTTTGTGCTGGAGGAGTATTTGCTGGATGCTCTCTTATTTGAATGGATCGCG 1054
QY 1021 aaggagtgatcactacatgggctaattctctgttacaatttcggtacagagatcttatg 1080
DB 1055 AAGGAGTGTGATCCTACATGGCTAATCTCTCTGTATACATTTTCGGTACGAGTATCTATG 1114
QY 1081 gcaagacactatttctcaagatactgctacccccaccaccacttcaagactcttat 1140
DB 1115 GCAGAGACACTATTTTTCAGATACTGCTGCTACCCAGCCACCATTTAAGACTCTTTAT 1174
QY 1141 tatacactcgtgattgattcttctgttaaggctctctccgggtgccttttctcgtctgttt 1200
DB 1175 TATACACTCGTGATTATGGATCTTTGTAAAGGCTCTTCGGGGTGCCTTTCTCTGCTGTGT 1234

/sub_species="japonica"
 /db_xref="taxon:39947"
 /clone="D40981"
 /dev_stage="shoot"
 76..2808
 /note="similar to Homo sapiens CBP80 and to rice product
 encoded by GenBank Accession Number D40981"
 /codon_start=1
 /product="nuclear cap-binding protein CBP80"
 /protein_id="AAG54079.1"
 /db_xref="GI:12621070"
 /translation="MSAGWRLRLRIGDRCEYGVSDHKEHICTCYGLVCLREYEHKS
 DAMEFLQLQADQLPHKIPFGLVGLINLENEDFSKGIQVDTTHANLQADLNENRDR
 IRLRLFLGLMSKVLPNSIIETFEALLSSAATILDEETGNPSQWPRADFYVVCIL
 ASLPMGSGELFEQVDEFEVLVGIQSVISIRRRHFDIAFSVFETDGRNSNPKDFIE
 DLWRIQVLSRNGKWKSVKPKPLHSFQAQLVAGVSHRFSPISCPHLHSNASSIEVKG
 QERHEADKLPORURRHIFPNKAENQPVDFVVEECILDVLLFNGCKEKAFLY
 VSLPFRYELMAETIFSQLLLPNPPFRPYITLVIIIDLCALPAFPFVYVGAVH
 ALFDRI SNMDECHTRLILWFSHHLSNFQFIWQWENAYVKDLKPWAPQRFVQEVLE
 REIRLSYFDKIQSIEDAVELEELLPKAGPNFRYHSDGKESDGHRLSKELVAMVR
 GRKTQGDILSWDEKIIIPVNGAKFALDVVSQTLIDIGSKSFHLITVLERGOIISKL
 CPNEEMQLLNDEVSATWKNSTQMIATIDRMGIRLLSNLAIKWFSPANVDQFHV
 SDRPWEILRLNVSINRIQDLRKEIQLRKGLOAKAESEKAAARELEFEAKSIIEVD
 GQVPSPNQLRLQADKAKGEVTEESLEAKALLARGLEESKELRLIFKSF
 VEVLRLPTISADQVNLNLRAGDQVNVSSARDPEATTMEIDNENGDNDSOLANGOK
 KISHNVELEQWCLDTGLYKLSFSROYATEIWSHIAMLDOEIFVGNTHPLIRKASPV
 CADLPVKGLFDEPDSSSELEFTCIIRKFFIMLSLTHGAQKYSVDYSGTSPILL"
 BASE COUNT 821 a 623 c 683 g 881 t 1 others
 ORIGIN

Query Match 36.0%; Score 978.2; DB 8; Length 3009;
 Best Local Similarity 62.5%; Pred. No. 3e-229;
 Matches 1642; Conservative 0; Mismatches 929; Indels 57; Gaps 5;

QY	46	tggaaacctctctctcgcacatcgcgaaaggagacctgagtcagcgacactccctccgac	105
DB	88	TGGAGGACGCTGCTGCTCGGATCGGCACAGGTCCCGGAGTAGCGGGGCTCGCTCGAC	147
QY	106	tacaaagaccacatcgagactgtttcgtgtcattcgtagagaaatcgagcttctgga	165
DB	148	CACAAGGAGCACATCGAAACTGTGTATGTGTGTCTTTGTCGACATACGAACTCCAAA	207
QY	166	gatcaagtttgcctttctactcaaatgctgaacattgcctcataagattcctttg	225
DB	208	GATCAATGTTGAGTTTCTCCTCAATGTGCAGATCAATTTGCCFCACAAATTCCTTTC	267
QY	226	tatgggactttgattggtttgtgaacttggagaatgaagattttgtccagaagctagta	285
DB	268	TTTGGAGTATTGATAGGTTTGATAAATTTGGAAATGAAGATTTTCCAAAGGTTATTGTC	327
QY	286	gaagtgctccacgtaatttccaggctgcttttagattcgtgcaactgcaacagtcacgt	345
DB	328	GATAACACATAGCAATTTACAGATGCTTGCATATGAAATCGTGACAGAAATCAGG	387
QY	346	atattgcttgcttatgacttccctgtgtgcagtaaggttatccaacctgcttcttg	405
DB	388	ATATTGCTGCGNTTCTCTGTGGCTGTATGTCAGCAAGTTGTCTCTCCCAAAATTTCTATT	447
QY	406	atttgcgtcttctgaacattgctatcatctgctgcactactgtggatgaagagaagga	465
DB	448	ATTGAACATTTGAGGCATATTATCATCTGCTGCAACAATATTAGATGAGGAACCCGGA	507
QY	466	aatccatcattggcagccacagctgacttttacgttatatgcatttgcagcctcccg	525
DB	508	AATCCTTCGTGGCAACACGCTGATTTCTATGTTTATTGTTATCTTGGCTTCCTTCCA	567
QY	526	tgggagagatcagaaactcgtcagcaagttcctgatgagatgaagaggtttagttggg	585
DB	568	TGGGGTGGCTCAGAAATTTGTTTGAGCAAGTTTCCAGATGAATTTGAGAGAGTCTGTGGT	627
QY	586	atacaagcttatttgacatccgaagaattcttccaccttgggttaaacctttttcac	645
DB	1702	GACATTGGCTCAAAAAGTTTCCACCATCTTATCACTGTTTGGAGAGATATGTCACAAATA	1761

DB	628	ATACAGTCTTATATAGCATTAGAAGGCATTTTGTATGATATTCCTTTTCAGTCTTTGAA	687
QY	646	aacggagaatttgaagcagccttgcagagaagatttcgtggagagatctatttggatcga	705
DB	688	ACAGATGAA--GGCAACTCTCCCAACAAAGAGATTTCATCGAAGATTATGGAGCGT	744
QY	706	attcagctctggtctccaatgatggaacttgaagcgtacctaggctcctactctcg	765
DB	745	ATTCAAGTCTTTCGCAANTGGTGAAGGTTAAGAGTGTCCAAACACTACCTGCTCG	804
QY	766	tttgaagctcaactcgttgcgaaagtttcatgagctacgtcccatataatgtatggaa	825
DB	805	TTTGAAGCTCAGCTGGTAGTGAGTTTCTCACCGTTTCTCCCAATATTAGTTGCCCCAC	864
QY	826	caacgagatccaccttctgatcatctcgaggggatacagtgggcaagcaagaagctatgca	885
DB	865	CTACTATCTCGCAATGCATCTCTGAAATAGTA--AAAGTCACAGAAAGCATGAAGCT	921
QY	886	ttgacgagatatccccagagaattcgtaggttgaaatatattccagctaaataaattggag	945
DB	922	GATCTGAAGTATCCCAAGGCTTCGTAGCTTCACATATTTCCAAACAATAAAGCTGAG	981
QY	946	gatgtacaaccaattgatcgttttgcgtggaggagatttgcgtgagtgctgcttctat	1005
DB	982	AACATGCAACCTGTAGATCGTTTGTGTGAAGAATGCATATTGGATGCTGCTACTTTC	1041
QY	1006	ttgaatgatgtcggaaagagtgtgcattcctacatggctaaattcttctgttaccattcgg	1065
DB	1042	TTCAATGGATGTCGCAAGAAATGTGCATTTTATCTGGTCAGCTTACCTGTGCTTCCGT	1101
QY	1066	tacgagtatcttatggcagagacactatttctcagatactgctgtaccccaagcacca	1125
DB	1102	TATGAATACCTGATGGCTGAGACCATATTTTCACAGCTACTGTTATTGCCGAATCCCT	1161
QY	1126	tccaagactctttatatcacactcgtgattgagctcttgcgttgaagctcttccgggtgcc	1185
DB	1162	TTCAGGCCAATTTACTATACCTTGGTTATTATCGACCTTTGCAAGCATTTGCCAGGTGA	1221
QY	1186	ttctcgt	1245
DB	1222	TTTCTCTCAGTTGTGTAGGAGCAGTACATGCTCTTTTGTGACAGAAATAGCAACATGGAT	1281
QY	1246	atggaatccagagacgctcttctcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1305
DB	1282	ATGGAGTCCCGACCCGACTTATCCTATGTTTTCACATCATCTGTGTTCAATTTTCAGTT	1341
QY	1306	atctggcgtgtgggaagtggt	1365
DB	1342	ATTGGCTTGGCAGAGTGGGCTTACGTCAGGACCTTCCAAATGGGCTCCACAGCGT	1401
QY	1366	gtatttgttggagagattttgcaagagaagtcagctgtgtgtgtgtgtgtgtgtgtgtgt	1425
DB	1402	GTTTTGTGCAAGAAAGTATTAGAAAGGGAATTCGCTTGTCTCTACTTTGACAAAAATAG	1461
QY	1426	cagagcattggaatcgcagactgcccctagaagaattacttctcccaaaagctgtgcgaat	1485
DB	1462	CAGACATAGAGGATGCTGTTGAATTTGGAAGAACTGTTTACCCCCAAAAGCTGGGCTTAAC	1521
QY	1486	tttatgtattccttggagaaggttaagagaaacagagaacagcaaatgtgacgcgaa	1545
DB	1522	TTGATATATCATGTGATGAAGGCAAGAAACACTGTATGGCCACAGACTCTCCAAAGGA	1581
QY	1546	ttgacgagaggttcaaggaacacacacccgcgtgacatgatagtgtggttgaagaa	1605
DB	1582	CTTGTGTCATGGTTAGAGGAAGGAGACACAAAGGTGATATTATTTCATGGGTAGACGA	1641
QY	1606	acgatatatacagttcagttgt	1665
DB	1642	AAAAAATTCCTGTAATGGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1701
QY	1666	gacatcgatcaaaaagtttccactcatcttggctcactgtccttgagcgaatatagccaagta	1725
DB	1702	GACATTGGCTCAAAAAGTTTCCACCATCTTATCACTGTTTGGAGAGATATGTCACAAATA	1761

CDS


```
RESULT 8
LOCUS DME238970 3228 bp mRNA linear INV 19-AUG-1999
DEFINITION Drosophila melanogaster mRNA for cap binding protein 80 (cbp80).
ACCESSION AJ238970
VERSION AJ238970.1 GI:5763574
KEYWORDS cap binding protein 80; cbp80 gene.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Lewis, J.D.
1 (bases 1 to 3228)
2 (bases 1 to 3228)
Unpublished
Direct Submission
Lewis, J.D.
Submitted (19-MAY-1999) Lewis J.D., Institute of Cell and Molecular
Biology, University of Edinburgh, Kings Buildings, Mayfield Road,
Edinburgh, Lothian, EH9 3JR, UNITED KINGDOM
FEATURES
source
1..3228
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
/map="90E1"
/tissue_type="imaginal disc"
/dev_stage="third instar"
269..2671
/gene="Cbp80"
269..2671
/gene="Cbp80"
/function="putative role in pre-mRNA splicing"
/codon_start=1
/product="cap binding protein 80"
/protein_id="CAB53186.1"
/db_xref="GI:5763575"
/translation="MSRRRAHDTDEGVDHRRNRRRVSENEIEDRLLESLLLRVEEA
TTSSVELEGVSVLEADLGTFLKILRLSDCAVAMPEKQVTTTLVGLLNKNYK
EDGTEVQVDRWFVFAVSLTPWVGRLDYKESALRELVNCHVISATSLQLLDMIDVSN
VSSDAPHPQEVLDCLNAQIRKLDQDNWAEKHIPRPVILFDSILCEALOHNLPTIV
PPHNFEPHPVYVYRFDYDCDGNLPGASIERFLIEEHLHIIETTHERKD
CAAQLSPYKHKKIPLEYIVVEVFAELFHMPTPRYLDICYGSIIECLKQPATLPQ
VLAQATEILFRIDSMTSCDFRVNFSYHLSNFKFTWSDEWDSCLLDGHEPRPK
FIOEVLQRLSYHORITEMPTTYAKLPLTPVPNYKYANEEANLPGTTVAHQLV
VAIROKCTPEEVNLIKIDIPNSGYSGEEMSGSNALKIDVFQTLNLGSKFSFSHF
AAISKFSVFEALATEEAQICILHNIPELWSSHQOMVVLIDKLQIVDCSAVAT
WIFSKEMGTFKLYLWELLHLITIKMKHVKIKLNTSEAEKAKADSSSDSEDD
SHRRKKPLTHADKPSEEVERMEKLEAANVNRQLFLIVFQRFIMILSEHLRLSDT
DGRDPDTPMYRTWITIGRLQQLVFLMHHEQVQKYSTLTETLFTSLDLDTHILEVQFVAL
RA"
BASE COUNT 845 a 807 c 807 g 769 t
ORIGIN
Query Match 2.4%; Score 65; DB 3; Length 3228;
Best Local Similarity 47.1%; Pred. No. 3.4e-05;
Matches 237; Conservative 0; Mismatches 260; Indels 6; Gaps 1;
QY 937 aaatggagaggtacacacaaatgatcgtttgtgctggagagagattgtgctggtg 996
|| || || || || || || || || || || || || || || || || || || || ||
Db 1181 AATCTCCGGGTCGCCACTAGAGCGCTTTTGTATTGAGGAGCATACACCATC 1240
|| || || || || || || || || || || || || || || || || || || || ||
QY 997 ctctctattgaatgagtgctggaagagtgctcctcacatgctgaatctcc--- 1052
|| || || || || || || || || || || || || || || || || || || || ||
Db 1241 ATTGAACATACCATCAGCGCGAAAGGATTGCGCCGCCAGCGTGTCTTTCGTAC 1300
|| || || || || || || || || || || || || || || || || || || || ||
QY 1053 --tggtacatttcggtacagtagtattctatggcagagacacattttctcagatactgctg 1110
|| || || || || || || || || || || || || || || || || || || || ||
```

```
/map="4C7-4C7"
1. 2975
/genes="Cbp80"
/notes="alignment with genomic scaffold AB003431"
/db_xref="FLYBASE:FBgn0022942"
256. 2658
/genes="Cbp80"
/notes="Longest ORF"
/codon_start=1
/db_product="FD31211p"
/protein_id="AAK93220.1"
/db_xref="GI:15291903"
/translation="MSRRADHDEGCHYHRRKRRVSENQIEDRLSILRVGR
STGSVENLGLVSLVLEADLGRFLKILRLSLDCAVRPEKCTVITLVGLLNKAKYK
FGGFVDMWKTFKESLKMCRMDAARSLRLADLVNCHVISATSLQLDITMDVSN
EDTPVQRMDFVAVLSTFWGRDLYEKESALELLRLIEVLYLNKRKKHNAIR
WASDAPHOEVLDCWLAQIRKLRODNMAEKHLPRLVLFVDSILCALOHNLPPIV
PPIHNDPEYPMWVYRMFDYTPCPDGNLPGAHISIERFLIEHLHI IETVHERKD
CAQQLSFPYKHKIPEYCIVEVVEAFELFHPMPYRLDICIYGSILIELKLOPATLQ
VLAQATEILPMRDSMTSDFRWNFSHLSNFKFTSWDEWSDCLLDGHEPRK
FIOELQKRLSHQRIEMPTTYAKLIPLPVPNYKYANEEANLPPTVAHQV
VAIROKTPPEVNAILDIPNSYSGEEMSDGSFNALKIDVFQTLNLGSKSHSF
AALSFKHSVFLAETEAQICILHNIIFELWSSHOQMMVLIDKLLOIVDCSAVAT
WIFSKMTGFTKLYLWEILHLITKMKHVIKINTLSEAKELAKLAKADSSSEDD
SSHKKRPITPHADKPSVEVVERMEKLEAANVQKRLFLIVFQRFIMLSEHLKSDT
DGRDPDWDIWRITGRLOQVFLMHHEQVQKYSSTLETLTSLDLDTHILEVFOQFVAL
RA"
BASE COUNT 765 a 772 c 768 g 670 t
ORIGIN

Query Match 2.38; Score 61.8; DB 3; Length 2975;
Best Local Similarity 46.4%; Pred. No. 0.00021;
Matches 241; Conservative 0; Mismatches 272; Indels 6; Gaps 1;

QY 937 aaatggaggatgtacaaacaaatgatctgttttgcgtggaggagatattctgagatgtg 996
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 AATCTGCGGGTGGCCACTCAATAGAGCGCTTTTGTATGAGGAGCATCTACACCATATC 1227
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 997 ctctctattgaatgatgtcggaagagtgatgcacatcactacatgccttaattcttc--- 1052
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 ATTGAACACATACCATCAGCAGCAAGAGTTCGCGCGCCAGCTGCTTAGCTTCCGATC 1287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1053 --tgtttacattcgtgaagatctctatgcagcagacactattttctcagatcgtg 1110
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 AAGCACAGATCCCGCTGGAGTACTGTCATCTGAGGTGGTCTTCGCGAGCTCTTTTCAT 1347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1111 ctacccagccaccattcaagaactctttattatacactcgtgattgattgattgtaag 1170
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 ATGCCCCAGCGCGCTATCTGGACATCTGCTACGCTCCATTTGATCGAACTCTGCAAA 1407
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1171 gctctccgggtgcttctctgctgtgtgtgctgctgtgtgtgtgtgtgtgtgtgtgtgt 1230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 CTGACGCGCGCCAGCTGCCCCAGGTTCTGCGCCAGGCCACCGAGATCTCTTATGGCG 1467
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1231 atatccgacttagacatggaatccagacgcgttattatctctgtgtgtgtgtgtgtgtgt 1290
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1468 ATGACACTGATGAACAGCTCATCTTTTCATCGGTTCGTCATATGGTTCTCTCATCATCTG 1527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1291 tccaacttcaattcattcgtccgtgggaagagtgggctttttgtgtgtgtgtgtgtgtgtgt 1350
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1528 AGCAACTTCAGTTTCATATGTCATGGACCAAGTGGACAGCTGCTCTCTCTCGACGGC 1587
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1351 tggggcccttaagcgtgtattttgttcaggagagattttgcaagagaagtcgctgtcttacc 1410
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1388 GAGCATCCGCTGCCAAATTCATCAGGAGGTGTTCGAGAGTGCTCGCTACGCTTTCGTAT 1647
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1411 tgggataaaattagcagagagcatttgaagaatcgactgcc 1449
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1648 CATCAGCGCATTCAGGAAATGATGCCAACGACTTATGCC 1686
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 10

```
HUMNCBP
LOCUS
DEFINITION Human mRNA for nuclear cap binding protein, complete cds.
ACCESSION D32002
VERSION D32002.1 GI:1865647
KEYWORDS nuclear cap binding protein.
SOURCE Homo sapiens
ORGANISM Homo sapiens
cell_line="HeLa"
cdna to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3066)
Kataoka,N.
Direct Submission
Submitted (29-JUN-1994) Naoyuki Kataoka, Faculty of Science, Kyoto
University, Department of Biophysics, Kitashirakawa-Oiwake,
Sakyo-ku, Kyoto 606, Japan
(E-mail: g53376@sakura.kudpc.kyoto-u.ac.jp, Tel: 075-753-4232,
Fax: 075-753-4235)
2 (bases 1 to 3066)
Kataoka,N., Ohno,M., Kangawa,K., Tokoro,Y. and Shimura,Y.
Cloning of a complementary DNA encoding an 80 kilodalton nuclear
cap binding protein
Nucleic Acids Res. 22 (19), 3861-3865 (1994)
95023141
MEDLINE
COMMENT On Mar 5, 1997 this sequence version replaced gi:577627.
FEATURES
Location/Qualifiers
1..3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HeLa"
265..2637
/codon_start=1
/product="nuclear cap binding protein"
/protein_id="BAA06769.1"
/db_xref="GI:577628"
/translation="MSRRHSDNDGGOPKRRKTSNDANETEDHLESICKVGEKSAK
SLESNGLAGVLEADLPNYKSKILRLCTVARLPKLTITTLVGLLNKAKYK
EFVAMIRQLKESLKMCRMDAARSLRLADLVNCHVISATSLQLDITMDVSN
VQVQRMDFVAVLSTFWGRDLYEKESALELLRLIEVLYLNKRKKHNAIR
ADKPHOEVLDCLWLAQIRKLRODNMAEKHLPRLVLFVDSILCALOHNLPPIV
TEDSVYPMRVIIFRMFDYTPCPDGNLPGAHISIERFLIEHLHI IETVHERKD
QVYSGNKKIPLNYHIVEVVEAFELFHPMPYRLDICIYGSILIELKLOPATLQ
QATENLYRLDTHMTCTVDRIINFSHLSNFKFTSWDEWSDCLSDQSPSPKPFVR
EVLKMRSLYHQRLIDLPVTFPSALCPANPTCIYKYGDESSNDSCLSPQSLPOVLA
KSKATNDEIFSLKDPNODDDDEGFSNPKIEVFQTLHLAAKSSHSFSAI
AKHFVEFKTLAESDEGKLHLVLRVFNVRNHPQMIATVDRKMTQIVDCAVANWIF
SSELSRDTFLRVFWELHSTIRKMKHVKLQKELEAKELAROKHRRSDDDDRSSD
RKDGVLSEQIERLQEKVESAGEOKNLFVIFQRFIMLSEHLKSDTGTGTSVLTPIWY
KNCIERQQIFLQHHQIQQIMVTLNLLFTAEILDPHILAVFQQFCAQA"
polyA_signal 3043..3048
polyA_site 3066
BASE COUNT 870 a 660 c 699 g 837 t
ORIGIN

Query Match 2.1%; Score 57.2; DB 9; Length 3066;
Best Local Similarity 51.1%; Pred. No. 0.0028;
Matches 134; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 1644 aatagttgtacagacctacttgacatcggatcggatcaaaaggtttcactcatttggtcactgt 1703
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1899 AGTCTTTGTACAGACTCTGCTACACTTGGCAGCCAAATCATTTACGCCACTCTCTTCTAGTC 1958
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1704 cctgagcgatgtgccaagtattttcaagctttgtcctgatacagataagcaggtgat 1763
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1959 TCTTGCAAGTTTACGAGGCTCTCAAAACCCCTAGCTGAAGTGTGATGAAGGAAGTTACA 2018
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1764 gctattatctcaagtgtacatactggaaaaaacaaatgtacaaatgacgcggtggcaat 1823
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2019 TGTGCTAAGACTTATGTTGAGTCTGGAGGAGCAACCATCCACAGATGATGCTGTACTACT 2078
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1824 tgataggatgatgggtttatagactagtatttataatcaggcaattgttagatgggtgtcttc 1883
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 20, 2002, 11:35:21 ; Search time 2808.13 Seconds
(without alignments)
13054.143 Million cell updates/sec

Title: US-09-882-986-1

Perfect score: 2716

Sequence: 1 aaagagacgaactgaagaaa.....ctttaaaaaaaaaaaaaa 2716

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estlin:*
- 4: em_estm:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_estc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	590	21.7	591	9 AV555064	AV555064 AV555064
C 2	529.2	19.5	535	9 AV555295	AV555295 AV555295
C 3	442.2	16.3	509	9 AV441539	AV441539 AV441539
C 4	417.6	15.4	424	10 BE523479	BE523479 M37F3STM
C 5	413.8	15.2	575	12 BH587474	BH587474 B0HN040TF
C 6	401	14.8	401	9 AV564181	AV564181 AV564181
7	383.8	14.1	480	10 R90218	R90218 16573 Lambd
8	347.8	12.8	516	9 AA605490	AA605490 30526 Lam
9	315.2	11.6	595	9 AW306466	AW306466 se51b03.y
10	311.8	11.5	682	10 BI935623	BI935623 EST55512
11	309.4	11.4	663	10 BI933835	BI933835 EST553724
12	294.2	10.8	643	10 BI935668	BI935668 EST555857
13	288.2	10.6	637	9 AW267982	AW267982 EST306204
14	273.2	10.1	466	9 AL365935	AL365935 M1BA03C07
15	267.8	9.9	632	10 BE590437	BE590437 SB177 Sug
16	258.6	9.5	585	9 AI894891	AI894891 EST264334
17	257	9.5	533	10 BI785886	BI785886 sa130c09.

18	251.2	9.2	593	10 BE402476	BE402476 CSB008002
19	246.4	9.1	552	10 BE492845	BE492845 WHE0565_F
20	243.8	9.0	584	10 BF619837	BF619837 HVSMEC000
21	240	8.8	240	10 BE530285	BE530285 M77G15STM
22	237.4	8.7	291	9 AV528587	AV528587 AV528587
23	232.4	8.6	371	9 AI442604	AI442604 sa43a09.y
24	226.2	8.3	481	9 AW906575	AW906575 EST342697
25	221	8.1	458	10 BI935752	BI935752 EST555641
26	201	7.4	242	9 AV560536	AV560536 AV560536
27	195.2	7.2	497	10 BG599218	BG599218 EST504118
28	191.6	7.1	490	9 AW568693	AW568693 sf60h04.y
29	174.6	6.4	895	10 BG368900	BG368900 HVSMEI002
30	174.2	6.4	466	10 D40981	D40981 RICS3204A.R
31	173.8	6.4	372	10 BE400529	BE400529 AWB004.E0
32	164.8	6.1	288	10 BI120970	BI120970 F025P93Y
33	157	5.8	585	10 BJ182040	BJ182040 BJ182040
34	150	5.5	356	9 AV409428	AV409428 AV409428
35	145.4	5.4	568	10 BJ188993	BJ188993 BJ188993
C 36	141.8	5.2	794	10 BE824117	BE824117 GM700023A
37	141.2	5.2	332	10 BE600932	BE600932 P11_92_H1
38	137	5.0	210	10 F15394	F15394 ATTS5642 Gr
39	134.2	4.9	394	10 D41074	D41074 RICS3340A.R
40	137.4	4.7	518	10 BF069198	BF069198 st43e07.y
41	125.6	4.6	276	10 BM444036	BM444036 EBem09_SO
42	123	4.5	424	10 C72488	C72488 C72488 Rice
43	118	4.3	526	10 BJ177704	BJ177704 BJ177704
C 44	115.2	4.2	714	9 AW348783	AW348783 GM210003A
45	108.6	4.0	602	10 BJ194498	BJ194498 BJ194498

ALIGNMENTS

RESULT 1	AV555064/c	591 bp	mrna	linear	EST 07-SEP-2000
LOCUS	AV555064	Arabidopsis thaliana	green siliques	Columbia	Arabidopsis
DEFINITION	thaliana	CDNA clone SQ003a06F 3',	mrna	sequence.	
ACCESSION	AV555064				
VERSION	AV555064.1	GI:8726479			
KEYWORDS	EST				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1 (bases 1 to 591)				
AUTHORS	Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.				
TITLE	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries				
JOURNAL	DNA Res. 7, 175-180 (2000)				
MEDLINE	20363093				
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.				
FEATURES	Location/Qualifiers				
source	1..591				
	/organism="Arabidopsis thaliana"				
	/strain="Columbia"				
	/db_xref="taxon:3702"				
	/clone="SQ003a06F"				
	/clone_lib="Arabidopsis thaliana green siliques Columbia"				
	/tissue_type="green siliques"				
	/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"				

BASE COUNT	161 a	138 c	117 g	175 t
ORIGIN				

Query Match 21.7%; Score 590; DB 9; Length 591;

Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers
1. 509

FEATURES

source
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AP263f01.f"
/clone_lib="Arabidopsis thaliana above-ground organ two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 141 a 109 c 104 g 155 t
ORIGIN

Query Match 16.3%; Score 442.2; DB 9; Length 509;
Best Local Similarity 93.3%; Pred. No. 1.7e-74;
Matches 47; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 2185 gagcaaaaggctctttaaagagctctcttgagacggaggtttttactgcttg 2244
|||||
DB 509 GAGCAAAAGAGGCTCTTTACAGAGCTCTCTTAAGCACACAGGTTTACAGCTCTTG 450
QY 2245 ctgttccaaagtcttcttaggtgtagtgaagacggctccagatcccaactaaagtga 2304
|||||
DB 449 CTGTCCCAAGTCTTCTTAGGTGTAATAAAGGATCGCTCCAGATCCCAACTAAAGTGAGA 390
QY 2305 tcagtgcagatctaaatctataggtgctgaagatgacaaagccatctgcgatggacgtg 2364
|||||
DB 389 TCAGTCGAGGATCTAAATCTATAGGTCTGAAGATGCAAGCCCATCTTCGATGGACGTG 330
QY 2365 gacgcgagaa-tggaaacccaaagaggtgcgaagtcggtgagagagaaacagtggtg 2423
|||||
DB 329 GACAGCGAGAAATGGAAACCCAAATAGAGTTCAAAAGTCGGTGAGAGAAACAGTGGTG 270
QY 2424 cttatcaacacttggtctatctcagcggcatttacaggccaatgtagcgagatggtcc 2483
DB 269 CTTATCAACACTTGGCTATCTCAGGGCATTTACAGGCAATATGCGCAAGATATGGCC 210
QY 2484 tcacatgagagtgtagtcagagtggtctctcggtgaagatggtcctctcttct 2543
|||||
DB 209 TCACATGGAGAAGTGGAGTCAAAAGTGTCTCGGTGAAGATGTCATCTCTCTTCT 150
QY 2544 ccaagccatatctctgcacttcaattcccaattacattatctctctcttcaactcaaa 2603
DB 149 CCAAGCCATATCTCTGCACATCAATTCCTCCATACATTAATCTCTTTCATCTCAA 90
QY 2604 tcaaacctgtctcttctgtttgtttgtatgagattctgattctgacatcaagttattagg 2663
DB 89 TCAACCTGTCTCTTCTTTTGTATGAGATTGGTTCTGGTCTGGCATCAAGATATTAGG 30
QY 2664 aaattgaaagagtcacaaaacagag 2690
|||||
DB 29 CAATTGAAAGATGTCAAAAGAAACAAG 3

RESULT 4
BE523479 424 bp mRNA linear EST 19-MAR-2001
LOCUS M37F3STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M37F3 5', mRNA sequence.
ACCESSION BE523479
VERSION BE523479.1 GI:9781457
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
TITLE A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil plant. Physiol. 124 (4), 1582-1594 (2000)
JOURNAL 20567808
MEDLINE
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142930603 TEL: 6142929371

FEATURES

source
1. 424
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M37F3"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 105 a 93 c 92 g 134 t
ORIGIN

Query Match 15.4%; Score 417.6; DB 10; Length 424;
Best Local Similarity 99.1%; Pred. No. 8.8e-70;
Matches 420; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 24 ctctcggaagaagatgagcaattggaaaaactctctcttcgcacatcgcggaagggacc 83
DB 1 CGCACGAGAGAAGATGAGCAATTGGAAACTCTCTCTTCGTCATCGCGCAAGGACG 60
QY 84 tgagtagcggcaactctctccgaactacaaagacacatcgagactgtttcggtgtcattcg 143
DB 61 TGAGTACGGCACTTCTCGGACTACAAAGACCACATCGAGACTTGTTCGGTGTCTATTG 120
QY 144 tagaagaatcgagcgtctctggagatcaagtcttgcctttctactacaatgtctgaaca 203
DB 121 TAGAAGAAATCGAGCGTCTCGGAGATCAAGTTTTCCTTTTCTACTACATGTCTGACCA 180
QY 204 attgcctcataagattctcttgggactttgattgattgttgaacttggagaatga 263
DB 181 ATTGCTCATAGATTCTTTGTATGGGACTTTGATTGTTTCTTCAACTTGAGAATGA 240
QY 264 agattttgtccagaagctagtagaagtgctccacgctaatttccaggctcgcttagttc 323
DB 241 AGATTTTGTCCAGAAGCTAGTAGAAGTGCTCCACGCTAATTTCCAGGTCTCTTAGATT 300
QY 324 tggcaactgcaacagatccgttatattgctgctttatgacttccctcttgcagtaa 383
DB 301 TGGCAACTGCAACAGTATCCGTATATTGCTTCCTTTATGACTTCCCTGTGTGACAGTAA 360
QY 384 ggtattcaacctgctctcttggattgctgttcgaaacattgctatcatctgctgccac 443
DB 361 GGTATTCAACCTGCTCTTTGATTGTCGCTCTCGAATTCGCTATCATCTCTGCCAC 420
QY 444 tact 447
DB 421 TACT 424
RESULT 5


```

Qy 1757 aggtgagtctattctcaagtgcgtacatactgtaaaac 1797
|||||
Db 361 AGGTGAGCTATTATCTCAAGTGAGTACATCTACTGGAAAAAC 401

RESULT 7
R90218
LOCUS 16573 Lambda-PRL2 Arabidopsis thaliana cDNA clone 191187, mRNA EST 30-DEC-1997
DEFINITION sequence.
ACCESSION R90218
VERSION R90218.1 GI:957758
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 480)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: T7 dye primer.
Location/Qualifiers
1..480
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="191187"
/clone.lib="Lambda-PRL2"
/notes="Vector: lambda Zip-Lox; Site.1: Sal; Site.2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."
BASE COUNT 124 a 80 c 125 g 132 t 19 others
ORIGIN
1..516
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="191187"
/clone.lib="Lambda-PRL2"
/notes="Vector: lambda Zip-Lox; Site.1: Sal; Site.2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."
BASE COUNT 126 a 114 c 118 g 138 t 20 others
ORIGIN

```


Query Match	11.6%	Score 315.2	DB 9	Length 595
Best Local Similarity	72.0%	Pred. No. 2.8e-50		
Matches 426	Conservative 0	Mismatches 163	Indels 3	Gaps 1
Qy 1376	aggagatttgcgaagaaagtagcctgtcttactgggataaaaatcaagcagagcattg	1435		
Db 6	AGGAGGTTTGGAGAGGAAGTTCGCTTGTCACTACTGGACAAAGTTTAAGCAGAGTATTG	65		
Qy 1436	agaatgcgaatgccttagaagaattactctctccaaagctggtccgaatttttgatt	1495		
Db 66	AAATGCACCTGGTTTAAAGAAATGTCTCTCCAAAGGTTGGGCCGAACCTTTAGTTTGG	125		
Qy 1496	ccttgggaaggttaagagaaacacagaagaaacagcaetgttcagcgcgaatttgagcagga	1555		
Db 126	GTGCAGAGAGATGATPAAG---AAGCAANTGNACATGTACTGTCGGACAGCTCAACAACA	182		
Qy 1556	aggtcaaggaaaaaacaacccgcgcgtgacatgatagtgtggattgaaagaacgatatact	1615		
Db 183	TGGTTAAAGGAAAGGCACCTGTCCGTGAAATAATCTCATGGATTGATGAAAGTGTGGTTC	242		
Qy 1616	cagttcaatgttttgaagttaactcttcaatagttgtacagaccttacttgacatcgat	1675		
Db 243	CAAAATATGGTCTAGAAGTTACCCCTCAGAGTGGTTGTACAAACTCTTCTCAANTATTGGAT	302		
Qy 1676	cataaagtttcaactcatttggctcactgtccctggagcgatatggccaagtattttccaagc	1735		
Db 303	CCAAGAGCTTCACATCTTGTATGACTCTCTTGGAGAGATATGGCCAAGCTTCCTGCAAAAT	362		
Qy 1736	tttgtctgataacgataaagcagggtgatgtatctatctcaagttagtacatactcggaaaa	1795		
Db 363	TATCTCTGCACCGAGTAAGCAAGTCATGTGTAGTGTGAAGTGAAGTCTTCTTGGGAAGA	422		
Qy 1796	acaatgtacaaaatgacgcgggtggcgaattgataggatgtaggggttataagacttagtacta	1855		
Db 423	GTAATACCCAGATGACAGCAATAGCTATTGACAGGATGATGGGTATTACAGCTTCTTCAA	482		
Qy 1956	atcaggaatattgttagatgggtgttctctccagaaaaattgtgatacagtttcaatgtgctg	1915		
Db 483	ATCTGGCTATTGTGAGATGGGTCTCTCTCAGCAGAGAATATTGAGCAATTTTCATATGTCAG	542		
Qy 1916	atacgcaatgggaatactattggcaatgctcttaacaagacttataaacgcat	1967		
Db 543	ATCTGGCTTGGAGATTCTTAGAATTCGATTAAGCAACACACACAATTCGTAT	594		

Query Match	11.6%	Score 315.2	DB 9	Length 595
Best Local Similarity	72.0%	Pred. No. 2.8e-50		
Matches 426	Conservative 0	Mismatches 163	Indels 3	Gaps 1
Qy 1376	aggagatttgcgaagaaagtagcctgtcttactgggataaaaatcaagcagagcattg	1435		
Db 6	AGGAGGTTTGGAGAGGAAGTTCGCTTGTCACTACTGGACAAAGTTTAAGCAGAGTATTG	65		
Qy 1436	agaatgcgaatgccttagaagaattactctctccaaagctggtccgaatttttgatt	1495		
Db 66	AAATGCACCTGGTTTAAAGAAATGTCTCTCCAAAGGTTGGGCCGAACCTTTAGTTTGG	125		
Qy 1496	ccttgggaaggttaagagaaacacagaagaaacagcaetgttcagcgcgaatttgagcagga	1555		
Db 126	GTGCAGAGAGATGATPAAG---AAGCAANTGNACATGTACTGTCGGACAGCTCAACAACA	182		
Qy 1556	aggtcaaggaaaaaacaacccgcgcgtgacatgatagtgtggattgaaagaacgatatact	1615		
Db 183	TGGTTAAAGGAAAGGCACCTGTCCGTGAAATAATCTCATGGATTGATGAAAGTGTGGTTC	242		
Qy 1616	cagttcaatgttttgaagttaactcttcaatagttgtacagaccttacttgacatcgat	1675		
Db 243	CAAAATATGGTCTAGAAGTTACCCCTCAGAGTGGTTGTACAAACTCTTCTCAANTATTGGAT	302		
Qy 1676	cataaagtttcaactcatttggctcactgtccctggagcgatatggccaagtattttccaagc	1735		
Db 303	CCAAGAGCTTCACATCTTGTATGACTCTCTTGGAGAGATATGGCCAAGCTTCCTGCAAAAT	362		
Qy 1736	tttgtctgataacgataaagcagggtgatgtatctatctcaagttagtacatactcggaaaa	1795		
Db 363	TATCTCTGCACCGAGTAAGCAAGTCATGTGTAGTGTGAAGTGAAGTCTTCTTGGGAAGA	422		
Qy 1796	acaatgtacaaaatgacgcgggtggcgaattgataggatgtaggggttataagacttagtacta	1855		
Db 423	GTAATACCCAGATGACAGCAATAGCTATTGACAGGATGATGGGTATTACAGCTTCTTCAA	482		
Qy 1956	atcaggaatattgttagatgggtgttctctccagaaaaattgtgatacagtttcaatgtgctg	1915		
Db 483	ATCTGGCTATTGTGAGATGGGTCTCTTCAGCAGAGAATATTGAGCAATTTTCATATGTCAG	542		
Qy 1916	atacgcaatgggaatactattggcaatgctcttaacaagacttataaacgcat	1967		
Db 543	ATCTGGCTTGGAGATTCTTAGAATTCGAGTAAGCAACACACACAATTCGTAT	594		

```

RESULT 10
BI935623
LOCUS      682 bp mRNA linear EST 18-OCT-2001
DEFINITION EST555512 tomato flower, anthesis Lycopersicon esculentum cDNA
clone CTOD23B17 5' end, mRNA sequence.
ACCESSION BI935623
VERSION    BI935623.1 GI:16250095
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE  1 (bases 1 to 682)
AUTHORS    van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
            Uterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
            Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
            Generation of ESTs from tomato flower tissue, anthesis (2001)
            Unpublished (2001)
            Contact: CUGI
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES   Location/Qualifiers
            source          1..682
                        /organism="Lycopersicon esculentum"
                        /cultivar="TA496"
                        /db_xref="taxon:4081"
                        /clone_lib="tomato flower, anthesis"
                        /tissue_type="flower"
                        /dev_stage="anthesis"
                        /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                        XhoI; supplier: Cornell University; sequencing: The
                        Institute for Genomic Research; Flower buds and flowers
                        were taken from greenhouse plants (4-8 wks old, TA496).
                        They were immediately frozen in liquid nitrogen and then
                        size-separated while remaining frozen."
BASE COUNT 196 a 121 c 171 g 194 t
ORIGIN
Query Match      11.5%; Score 311.8; DB 10; Length 682;
Best Local Similarity 68.9%; Pred. No. 1.2e-49;
Matches 457; Conservative 0; Mismatches 202; Indels 4; Gaps 2;

QY 1445 ctgcccagaagaattacttctctccaaagtggtcgaaattttatgtattctcttggaag 1504
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CTGCTTTGGAGGAGCTACTTCCACCAAGAGTGGACCAACAATTTAAATATAGTGGCGAAG 60

QY 1505 aagtaagaagaacagaagaacgaatgttcagcgaattgagcaggaggtcaagg 1564
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATGTACTGTATCCAACTGAG---CGTGACACTTTCTTTAGAACTCAAGGACATGTGTAAAG 117

QY 1565 aaaaacaaacgcgcgtgacatagtggtgattgagaacacgatatatccaggtcatg 1624
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 GTAGAAACATGCGCGGAATGATCTCTCGGTAGAGAAANAATGTTTTCCAGCCCATG 177

QY 1625 gtttgaagtactcttacaatagttgtacagaccttacttgacatcgatgatacaaaagt 1684
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 GGTTTGATATTACCTTGGAGTTGTTGTTCAACACACTTCTCGACATCGATCAAAAAGCT 237

QY 1685 tcaatctttgttcactgctctcgagcgatgagccaaagtattttcaagctttgtctctg 1744
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TTACTCATTTGATTACTGTCTTGGAGAGATATGCCAAGTATTGCCAAGATGTGTACTG 297

QY 1745 ataacgataagcaggtgacttatcttcaagtgaagtacatactggaacaaatgtac 1804
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 298 ATGATGATCAACAAGCTTAAGCTGATTACGGAAGTAAGTTCTTACTGCGAAAAACAGTGGCG 357
QY 1805 aaatgaacggcgggtgcaattgatagatgattggttatagactagtagtatcatcaggca 1864
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 AAATCACAGCTATTGCCATCGACCGGATGATGAGTTATCGGCTTATATCTAATTTGGCCA 417
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1865 ttgttagatgggtgttctctccagaaaaatgttgatcagttttcatgtgtctgacagccat 1924
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TAGTGAGATGGGTTTCTCTCCACTGAATCTTGATCGGTTTCATGTTCTGACAGTTCGT 477
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1925 gggagatacttggaatgctctttaaacaagactataaacgatatctctgtattgaggaag 1984
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 GGGAGATCCTCAGGAATGCTGTCAGTAAGACGCTATAATCGGATTTCTGTATCTGAGAAAAG 537
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1985 atatatcaaacattacgaaaaaattgtttggttggtgagaaaaagcttcagccaatgcacgg 2044
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 AGATTTCATCCTCTGAGAGAAAGTGTGTCTTGGCTGA-AAAGCTGCATCTAGAGCCAGAG 596
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2045 tagagttgagggctgtagagcaaaccttccctagtggaagtgaaacccgttcttgggtg 2104
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 AAGATTAGAAATCTCCGAGTCTAAGCTGAGTGTATAGATGGGAACCTGTTCTTTGGTG 656
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2105 aga 2107
|||
Db 657 AAA 659

RESULT 11
BI933835
LOCUS      663 bp mRNA linear EST 18-OCT-2001
DEFINITION EST553724 tomato flower, anthesis Lycopersicon esculentum cDNA
clone CTOD18E1 5' end, mRNA sequence.
ACCESSION BI933835
VERSION    BI933835.1 GI:16248307
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE  1 (bases 1 to 663)
AUTHORS    van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
            Uterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
            Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
            Generation of ESTs from tomato flower tissue, anthesis (2001)
            Unpublished (2001)
            Contact: CUGI
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES   Location/Qualifiers
            source          1..663
                        /organism="Lycopersicon esculentum"
                        /cultivar="TA496"
                        /db_xref="taxon:4081"
                        /clone_lib="tomato flower, anthesis"
                        /tissue_type="flower"
                        /dev_stage="anthesis"
                        /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                        XhoI; supplier: Cornell University; sequencing: The
                        Institute for Genomic Research; Flower buds and flowers
                        were taken from greenhouse plants (4-8 wks old, TA496).
                        They were immediately frozen in liquid nitrogen and then
                        size-separated while remaining frozen."
BASE COUNT 186 a 120 c 167 g 190 t
ORIGIN

```

Query Match	11.4%;	Score 309.4;	DB 10;	Length 663;
Best Local Similarity	68.5%;	Pred. No. 3.5e-49;		
Matches 457;	Conservative 0;	Mismatches 206;	Indels 4;	Gaps 2;
<hr/>				
QY 1445	ctgccctagaagaattactctcccaaaagctggtccgaattttatgtattcttctggaag	1504		
Db 1	CTGCTTTGGAGGAGCTACTTCCACCAAGAGTGGACCACAATTAAATATAGTGGCGAAG	60		
<hr/>				
QY 1505	aaggttaagagaagaacagacagaatgtccagccgaattgagcaggaagtccaag	1564		
Db 61	ATGGTACTGATCCAACTGAG---CGTGCACATTTCTTTAGAACTCAAGGACATGGTGAAG	117		
<hr/>				
QY 1565	aaaaaacaacccagcgtgacatagtggtgattgaagaaacgatatatccagtctatg	1624		
Db 118	GTAGGAAAACTCGCCGGAATGATCTCTCGGTAGNAGAAATGTTTTCAGGCCCATG	177		
<hr/>				
QY 1625	gttttgaagtactctctacaaatagttgtacagaccttacttgacatcgatgatacaaaagtt	1684		
Db 178	GGTTTGATATTACCTTTGGAGTGTGTTTCAAAACACTTCTCGACATCGGATCAAAAAGCT	237		
<hr/>				
QY 1685	tcactcatttggctcgtcctgagcgatagtgccaagtattttcaaaagcttctgctg	1744		
Db 238	TTACTCATTTGATCTGCTTGGAGAGATATGCCCAAGTCATGTCAAAGATGTGTACTG	297		
<hr/>				
QY 1745	atacgaataagcaggtgatgctattatctcaagtgaatcactggaatacaaatgtac	1804		
Db 298	ATGATGATCAACAAGTTAAAGCTGATTAACGGAAGTAAGTTCTTACTGGCAAAACAGTGGC	357		
<hr/>				
QY 1805	aaatgacggcggtyggcaattgatatagtagtgatgggttatagactagtagtaatacaggcaa	1864		
Db 358	AAATGACAGCTATTGCCATCGCGGATGATGATGTTTCAAAACACTTCTCGACATCGGATCAAAAAGCT	417		
<hr/>				
QY 1865	ttgttagatgggtgtctctccagaaaaatgttgaatcagtttcatgtctgtatcagccat	1924		
Db 418	TAGTGAGATGGGTTTCTCTCCACTGAAATCTTGATCGGTTTCATGTTTCTGACAGTTGCT	477		
<hr/>				
QY 1925	ggagataacttggcaatgctcttaacaagactataaccgtatctctgtattgtggaag	1984		
Db 478	GGGAGATCCTCAGGAATGCTCTCAGTAAGAGATATATCGGATTTCTGATCTGAGAAAAG	537		
<hr/>				
QY 1985	atatatacaacattacgaaaaattgtttgttctgagaaagcttccagcaatgcacag	2044		
Db 538	AGATTTTCATCCCTTGAGAACAGTGTGCTCTTGCTGA-AAAGCTGCATCTAGAGCCAGAG	596		
<hr/>				
QY 2045	tagagttggaagctgctgagagcaaaactttccctagtgggaagtgaaacccgttcttgggtg	2104		
Db 597	AAGAGTTAGAAATCTGCCGAGTCTAAGCTGAGTGTCTAGATGGGGAACCTGTTCTTGGGT	656		
<hr/>				
QY 2105	agaatcc 2111			
Db 657	GAACCC 663			
<hr/>				
RESULT 12				
BI915968				
LOCUS	BI915968	643 bp	mRNA	linear
DEFINITION	EST353857 tomato flower, anthesis Lycopersicon esculentum cDNA clone cTOD24H3 5' end, mRNA sequence.			
ACCESSION	BI915968			
VERSION	BI915968.1	GI:16250440		
KEYWORDS	EST.			
SOURCE	tomato.			
ORGANISM	Lycopersicon esculentum			
REFERENCE	1 (bases 1 to 643)			
AUTHORS	van der Hoeven,R.S., Bezzerides,J.I., Karamycheva,S.A., Tsai,J., Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.			
TITLE	Generation of ESTs from tomato flower tissue, anthesis (2001)			

Unpublished (2001)
Contact: CUCI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics Institute

Seq primer: T3.
Location/Qualifiers
1. 643
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOD24H3"
/clone_lib="tomato flower, anthesis"
/tissue_type="flower"
/dev_stage="anthesis"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Cornell University; sequencing: The Institute for Genomic Research; flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."
BASE COUNT 183 a 114 c 163 g 183 t
ORIGIN

Query Match 10.8%; Score 294.2; DB 10; Length 643;
Best Local Similarity 68.0%; Pred. No. 2.9e-46;
Matches 440; Conservative 0; Mismatches 203; Indels 4; Gaps 2;

QY 1445 ctgccctagaagaattactctcccaaaagctggtccgaattttatgtattcttctggaag 1504
Db 1 CTGCTTTGGAGGAGCTACTTCCACCAAGAGTGGACCACAATTAAATATAGTGGCGAAG 60
QY 1505 aaggttaagagaagaacagacagaatgttcagccgaattgagcaggaagtccaag 1564
Db 61 ATGGTACTGATCCAACTGAG---CGTGCACATTTCTTTAGAACTCAAGGACATGGTGAAG 117
QY 1565 aaaaaacaacccagcgtgacatagtggtgattgaagaaacgatatataccagtcttctg 1624
Db 118 GTAGGAAAACTCGCCGGAATGATCTCTCGGTGAGAAATAATGTTTTCAGGCCCATG 177
QY 1625 gttttgaagtactcttacaatagttgtacagaccttacttgacatcgatgatacaaaagtt 1684
Db 178 GGTTTGATATTACCTTTGGAGTGTGTTTCAAAACACTTCTCGACATCGGATCAAAAAGCT 237
QY 1685 tcactcatttggctcgtcctgagcgatagtgccaagtattttcaaaagcttctgctg 1744
Db 238 TTACTCATTTGATCTGCTTGGAGAGATATGCCCAAGTCATGTCAAAGATGTGTACTG 297
QY 1745 atacgaataagcaggtgatgctattatctcaagtgaatcactggaatacaaatgtac 1804
Db 298 ATGATGATCAACAAGTTAAAGCTGATTAACGGAAGTAAGTTCTTACTGGCAAAACAGTGGC 357
QY 1805 aaatgacggcggtyggcaattgatatagtagtgatgggttatagactagtagtaatacaggcaa 1864
Db 358 AAATGACAGCTATTGCCATCGCGGATGATGATGTTTCAAAACACTTCTCGACATCGGATCAAAAAGCT 417
QY 1865 ttgttagatgggtgtctctccagaaaaatgttgaatcagtttcatgtctgtatcagccat 1924
Db 418 TAGTGAGATGGGTTTCTCTCCACTGAAATCTTGATCGGTTTCATGTTTCTGACAGTTGCT 477
QY 1925 ggagataacttggcaatgctcttaacaagactataaccgtatctctgtattgtggaag 1984
Db 478 GGGAGATCCTCAGGAATGCTCTCAGTAAGAGATATATCGGATTTCTGATCTGAGAAAAG 537
QY 1985 atatacaacattacgaaaaattgtttgttctgagaaagcttccagcaatgcacag 2044
Db 538 AGATTTTCATCCCTTGAGAAAAGTGTGCTCTTGGCTGA-AAAGCTGCATCTAGAGCCAGAG 596
QY 2045 tagagttggaagctgctgagagcaaaactttccctagtgggaagtgaaacccgttcttgggtg 2104
Db 597 AAGAGTTAGAAATCTGCCGAGTCTAAGCTGAGTGTCTAGATGGGGAACCTGTTCTTGGGT 656
QY 2105 agaatcc 2111
Db 657 GAAACCC 663

Search completed: August 20, 2002, 12:28:33
Job time: 3192 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:39:56 ; Search time 105 98 Seconds
(without alignments)
6294.967 Million cell updates/sec

Title: US-09-882-986-1

Perfect score: 2716

Sequence: 1 aaagagcgaactgaagaaa.....ctttaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	45.2	1.7	7218	1	US-08-232-463-14
2	45	1.7	1386	2	US-08-897-340-1
3	45	1.7	1386	3	US-09-252-329-1
4	42.8	1.6	7218	1	US-08-232-463-14
5	42.4	1.6	276	4	US-09-328-111-820
6	42.2	1.6	1248	4	US-09-020-956-171
7	42.2	1.6	1248	4	US-09-030-607-171
8	42.2	1.6	1248	4	US-09-439-313-171
9	40.8	1.5	12720	1	US-08-403-866-11
10	40.2	1.5	3138	1	US-07-867-106-4
11	40	1.5	1214	4	US-09-439-313-225
12	40	1.5	1215	4	US-09-439-313-326
13	39.2	1.4	642	3	US-08-946-026-33
14	39	1.4	4500	2	US-08-743-637B-35
15	39	1.4	4500	3	US-08-526-840B-35
16	39	1.4	5000	3	US-09-104-070-3
17	38.8	1.4	5852	1	US-07-867-106-2
18	38.8	1.4	15766	4	US-09-338-907-73
19	38.8	1.4	15766	4	US-09-218-207-73
20	38.8	1.4	37950	4	US-09-338-907-183
21	38.8	1.4	37950	4	US-09-218-207-183
22	38	1.4	1751	2	US-08-953-138-3
23	37.4	1.4	358	2	US-08-721-488-9
24	37.4	1.4	746	4	US-09-013-810-1
25	37.2	1.4	1098	3	US-09-248-335-35
26	36.8	1.4	547	1	US-08-131-365B-41
27	36.8	1.4	547	2	US-08-668-123-41

28 36.8 1.4 11517 1 US-07-920-281C-1
29 36.8 1.4 11517 4 US-08-466-277-1
30 36.4 1.3 1117 4 US-09-247-373B-33
31 36.4 1.3 2407 4 US-09-370-807-7
32 36.4 1.3 4450 3 US-08-617-860B-2
33 36.2 1.3 2550 6 5258287-23
34 35.8 1.3 958 2 US-08-757-046A-5
35 35.8 1.3 958 3 US-09-447-208-5
36 35.8 1.3 958 3 US-09-135-988-5
37 35.8 1.3 958 4 US-09-277-716-5
38 35.8 1.3 958 4 US-08-597-274A-5
39 35.8 1.3 2852 3 US-09-027-137-2
40 35.8 1.3 4880 3 US-09-031-563-1
41 35.8 1.3 5125 3 US-09-031-563-4
42 35.6 1.3 1166 5 PCT-US96-12129B-1
43 35.6 1.3 1582 3 US-08-545-196B-10
44 35.6 1.3 1582 3 US-08-545-196B-12
45 35.6 1.3 3214 1 US-08-484-105-17

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Sequence 1, Appl
Sequence 1, Appl
Sequence 33, Appl
Sequence 7, Appl
Sequence 2, Appl
Patent No. 5258287
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 17, Appl


```

RESULT 7
US-09-030-607-171
: Sequence 171, Application US/09030607
: Patent No. 6262345
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FOR
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
:

```


US-08-946-026-33

Query Match 1.4%; Score 39.2; DB 3; Length 642;
Best Local Similarity 57.8%; Pred. No. 0.056;
Matches 89; Conservative 0; Mismatches 63; Indels

[illegible]

```

RESULT 14
US-08-743-637B-35/c
; Sequence 35, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273

```

Query Match

[illegible]

RESULT 15
US-08-526-840B-35/c
; Sequence 35, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: QUELETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENE
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995

Query Match 1.4%; Score 39; DB 3; Length 4500;
Best Local Similarity 53.6%; Pred. No. 0.23;
Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 2522 aagatgtgcacacctctctcttctccaaagccatatctctctgcacttcaattcccatattacatt 2581

Db 825 AAGTAGTTCACACTGTGCTTTTTTCTCTGCTTTTTTGTCTTGAATGCCTTAAATATCATATT 766
Qy 2582 aatcttcctcttcaatctcaatcaaacctgtctctctttgtttttgtttgttatgagattctg 2641
Db 765 ATTCCCTTTTCTAAGCTTCAATCCCTTCTCTATCTGTATTTCGCATAGCTACTACT 706
Qy 2642 attctgacatcaagttattaggaaaattgaaa 2672
Db 705 ATTTTACAGCAATCTATCTTTAAATATACA 675

Search completed: August 20, 2002, 13:51:46
Job time: 7910 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 11:41:01 ; Search time 413.38 Seconds

(without alignments)
11280.516 Million cell updates/sec

Title: US-09-882-986-1

Perfect score: 2716

Sequence: 1 aaagagagaactgaagaaa.....ctttaaaaaaaaaaaaaa 2716

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61.8	2.3	3224	ABL12159	Drosophila melanog
2	56.8	2.1	2902	ABL27678	Drosophila melanog
3	50.8	1.9	2244	ABL13529	Drosophila melanog
4	50.8	1.9	4420	ABL13528	Human immune syste
5	45.4	1.7	6242	ABL34148	Homo sapiens Tub I
6	45	1.7	1386	AAV11855	Tumour suppressor
7	44.4	1.6	6552	AA346678	Tumour suppressor
8	44	1.6	6189	AA346599	Human metastasis a
9	44	1.6	6189	ABL34592	

C	10	43.6	1.6	669	22	AA115089	Human breast cancer
	11	43.4	1.6	2872	13	AAQ28449	Cholesterol 7-alpha
	12	43	1.6	1456	22	AAD03657	Human serine prote
C	13	42.8	1.6	256	22	AAS07790	Cervical cancer pr
	14	42.8	1.6	612	22	AAH71471	Human cervical can
C	15	42.8	1.6	612	22	AAH71471	Human cervical can
C	16	42.6	1.6	396	22	AAF94862	Human ovarian can
C	17	42.6	1.6	425	22	AAS60450	Human cancer agent
	18	42.6	1.6	464	22	AAH33241	Human colon cancer
	19	42.6	1.6	3084	22	AAS25910	Human CDNA encodin
C	20	42.4	1.6	268	22	AAH70080	Human cervical can
C	21	42.4	1.6	276	21	AAZ80736	Human colon cancer
	22	42.4	1.6	411	22	AAI88595	Human polynucleoti
	23	42.2	1.6	1248	19	AAV61249	CDNA sequence of p
	24	42.2	1.6	1248	19	AAV58644	Prostate tumour sp
	25	42.2	1.6	1248	21	AAA06407	Human immunogenic
	26	42.2	1.6	1248	22	AAS63615	Human prostate CDN
	27	42.2	1.6	1248	22	AAI0166	Human prostate tum
	28	42.2	1.6	1248	22	AAH93523	Human prostate-spe
	29	42.2	1.6	1248	22	AAH84837	Human prostate-spe
	30	42.2	1.6	1248	22	AAH02588	Prostate tumour an
C	31	42	1.5	165	22	AAI84674	Human polynucleoti
	32	41.8	1.5	198	22	AAS07727	Cervical cancer pr
	33	41.8	1.5	1767	22	AAS29126	CDNA encoding for
	34	41.6	1.5	176	22	AAH25484	Nucleotide fragmen
	35	41.6	1.5	398	22	AAI88388	Human polynucleoti
C	36	41.6	1.5	534	22	AAS60058	Human cancer agent
	37	41.6	1.5	2498	21	AAA27988	Wheat CCR4 transcr
	38	41.6	1.5	4091	22	AAI58457	Human polynucleoti
	39	41.4	1.5	230	22	AAH33118	Human colon cancer
C	40	41.4	1.5	278	22	AAH89999	Human cervical can
	41	41.4	1.5	773	22	AAH64909	Human secreted pro
	42	41.4	1.5	930	22	AAH38866	Human colon cancer
C	43	41.4	1.5	1664976	19	AAV21209	Methanococcus jann
	44	41.2	1.5	2106	22	AAF95980	Nucleotide sequenc
	45	41.2	1.5	2907	21	AAC98917	Human pancreatic c

ALIGNMENTS

RESULT 1

ID ABL12159 standard; cDNA; 3224 BP.

XX ABL12159;

XX AC

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 30959.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB68056.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PS Claim 1; Page 1358; 3695pp; English.

The invention relates to human breast cancer expressed polynucleotides (AA07544-AA076789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

Sequence 669 BP; 260 A; 81 C; 74 G; 166 T; 88 other:

Query Match 1.6%; Score 43.6; DB 22; Length 669;
Best Local Similarity 47.3%; Pred. No. 0.16;
Matches 88; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy	2530	catcctctcttctccaaagccatctctctgcaactcaatcccatcataatctctcc	2589
Db	270	CNTCTTTTTTTTTTTNNTGTGNTNTATNAAAAAATTTNTTTTTTTTTTTNTTN	211
Qy	2590	tctttcaactcgaatcaaacctgctctcttctgttttttggatcgagatctgctcgac	2649
Db	210	NNTNGNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNGGCGCTNTCNA	151
Qy	2650	atcaagttattaggaaattgaaagagtcgaacaaacagaggttcaaacctttcaaaaaaaa	2709
Db	150	ATTAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	91
Qy	2710	aaaaaa 2715	
Db	90	AAAAAA 85	

RESULT	11
AAQ25449	
ID	AAQ25449 standard; DNA; 2872 BP.
XX	
AC	AAQ25449;
XX	
DT	07-DEC-1992 (first entry)
XX	
DE	Cholesterol 7-alpha-hydrolase gene.
XX	
KW	C7H; human; 7-hydroxycholesterol; recombinant; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	40..504
FT	/tag= a
XX	
PN	JP04144680-A.
XX	
PD	19-MAY-1992.
XX	
PF	05-SEP-1990; 90JP-0268587.
XX	
PR	05-SEP-1990; 90JP-0268587.
XX	
PA	(TOXN) TOYO JOZO KK.
XX	
DR	WPI; 1992-214120/26.
DR	P-PADB; AAR24482.
XX	
PT	Cholesterol 7-alpha-hydroxylase enzyme - and DNA sequence etc., for efficient prodn. of enzyme
XX	
PS	Claim 1; Fig 1; 15pp; Japanese.
XX	
CC	The DNA sequence encodes human cholesterol 7-alpha-hydrolase which

CC	catalyses the formation of 7-alpha-hydroxycholesterol, NADP and
CC	water from cholesterol, reduced NADP and molecular state oxygen.
CC	The enzyme may be produced by recombinant techniques by culturing
CC	cells transformed with a plasmid contg. the DNA sequence and
CC	recovering the enzyme from the culture. C7H may be produced
CC	efficiently by this method.
XX	
SQ	Sequence 2872 BP; 898 A; 532 C; 561 G; 881 T; 0 other;
Query Match 1.6%; Score 43.4; DB 13; Length 2872;	
Best Local Similarity 53.9%; Pred. No. 0.38;	
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;	
QY	2552 tatctctgcacttcaattcccatatcaattatctctctcttccaatccaatcaaacct 2611
DB	
QY	2704 tatcagctctatgccaagtatgatctctctgtagtattattataattattacct 2763
DB	
QY	2612 gtctctttgtttttgttgagattctgtattctgacatcaagttattaggaaattgaa 2671
DB	
QY	2672 aagagtcacaaacacagagtttaaacctttaaaaaaaaaaaaaa 2716
DB	
QY	2824 catctttcagaataaagtatgaatcctcaaaaaaaaaaaaaa 2868
RESULT 12	
AAD03657	
ID	AAD03657 standard; DNA; 1456 BP.
XX	
AC	AAD03657;
XX	
DT	19-JUN-2001 (first entry)
XX	
DE	Human serine protease, prost 07 DNA.
XX	
KW	Human; serine protease; PROST 07; gene therapy; vaccine; cytostatic;
KW	tumour metastasis; ribozyme therapy; antisense therapy; cancer; prostate;
KW	ovary; endometrium; benign prostatic hyperplasia; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	CDS 1..705
FT	/tag- a
FT	/product= "Human serine protease, PROST 07"
FT	1..30
FT	/tag- b
FT	/note= "Encodes pro-sequence and part of a signal
FT	sequence of human serine protease, PROST 07"
FT	31..703
FT	mat_peptide
FT	/tag- c
FT	/label= "Mature human serine protease, PROST 07"
XX	
PN	WO200125446-A1.
XX	
PD	12-APR-2001.
XX	
PF	05-OCT-2000; 2000WO-US27431.
XX	
PR	07-OCT-1999; 99US-0158588.
PR	02-OCT-2000; 2000US-0678940.
XX	
XX	(SCHD) SCHERING AG.
XX	
PI	Bringmann PW, Brink J, Harkins R, Light DR, Lin RJ, Parkes D;
PI	Parry G, Schneider DW, Steinbrecher R, Van Heult PT, Xuan J;
XX	
DR	WPI; 2001-266314/27.
DR	P-PSDB; AAE00397.
XX	
PT	Novel serine protease polypeptide and polynucleotides, designated PROST

	Query Match	1.6%	Score 43.4	DB 13	Length 2872
	Best Local Similarity	53.9%	Pred. No. 0.38		
	Matches 89	Conservative 0	Mismatches 76	Indels 0	Gaps 0
Qy	2552	tatctctgcacttcaatcccatatcaatctctctcttccaatctcaatcaaacct	2611		
Db	2704	tatcagctctcatgccaattgatattctctgctgagtatttatattaaatatattacct	2763		
Qy	2612	gtctctttgtttttgtttgagatctctgattctgacatcaagttattaggaaattgaa	2671		
Db	2764	ttttgctattttcacaaatattaagaanaattctctacatttgataattttgaaatgatt	2823		
Qy	2672	aagagtcacaaaaaacagaggttttaaacctttaaaaaataaaaaa	2716		
Db	2824	catctttcacaaataaaagtatgaattctaaataaaaaaataaaaaa	2868		

RESULT	12
AAD03657	
ID	AAD03657 standard; DNA; 1456 BP.
XX	
AC	AAD03657;
XX	
DT	19-JUN-2001 (first entry)
XX	
DE	Human serine protease, prost 07 DNA.
XX	
KW	Human: serine protease; PROST 07; gene therapy; vaccine; cytostatic;
KW	tumour metastasis; ribozyme therapy; antisenase therapy; cancer; prostate;
KW	ovary; endometrium; benign prostatic hyperplasia; ds.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	1..705
CDS	/*tag= a
FT	/product= "Human serine protease, PROST 07"
FT	1..30
sig_peptide	/*tag= b
FT	/note= "Encodes pro-sequence and part of a signal
FT	sequence of human serine protease, PROST 07"
FT	31..703
mat_peptide	/*tag= c
FT	/label= "Mature human serine protease, PROST 07"
XX	
WO	200125446-A1.
XX	
PN	
XX	
XX	
PD	12-APR-2001.
XX	
PF	05-OCT-2000; 2000WO-US27431.
XX	
PR	07-OCT-1999; 99US-0158588.
PR	02-OCT-2000; 2000US-0678940.
XX	
PA	(SCHD) SCHERING AG.
XX	
PI	Bringmann PW, Brink J, Harkins R, Light DR, Lin RJ, Parkes D;
PI	Parry G, Schneider DW, Steinbrecher R, Van Heult PT, Xuan J;
XX	
DR	WPI; 2001-266314/27.
DR	P-PSDB; AAE00397.
XX	
PT	Novel serine protease polypeptide and polynucleotides, designated PROT

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 15:25:57 ; Search time 87.13 Seconds
(without alignments)
1683.688 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLRLRIGKGPYGT.....EDVHPLFLQAISALQFPLH 848

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_nhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4374	100.0	848	10 Q9LKN6	Q9lkn6 arabidopsis
2	3784	86.5	749	10 Q9SIU2	Q9siu2 arabidopsis
3	2753.5	63.0	910	10 Q9AWB3	Q9awb3 oryza sativ
4	818.5	18.7	800	5 Q9U980	Q9u980 drosophila
5	818.5	18.7	800	5 Q9W412	Q9w412 drosophila
6	836.5	14.6	754	5 Q9I763	Q9i763 caenorhabdi
7	550.5	12.6	747	5 Q9VDA4	Q9vda4 drosophila
8	501.5	11.5	829	3 Q9HE91	Q9he91 neurospora
9	475.5	10.9	780	3 Q14253	Q14253 schizosacch
10	374.5	8.6	886	3 Q9Y717	Q9y717 candida alb
11	163	3.7	1413	5 Q9XWR0	Q9xwr0 caenorhabdi
12	145.5	3.3	1833	5 Q9VM67	Q9vm67 drosophila
13	145	3.3	1023	16 Q92B64	Q92b64 listeria in
14	142.5	3.3	802	4 Q96JN1	Q96jn1 homo sapien
15	142	3.2	2473	11 Q9Q284	Q9q284 mus musculu
16	140	3.2	906	4 Q96ST0	Q96st0 homo sapien

17	139.5	3.2	1937	13 Q9IBD4	Q9ibd4 gallus gall
18	139.5	3.2	3259	4 Q14789	Q14789 homo sapien
19	139	3.2	792	10 Q9AQO9	Q9aqo9 oryza sativ
20	139	3.2	1388	13 Q91785	Q91785 xenopus lae
21	138.5	3.2	3868	5 Q9W090	Q9w090 drosophila
22	138	3.2	1295	5 Q22257	Q22257 caenorhabdi
23	138	3.2	1930	13 Q9DGB5	Q9dgd5 pennahia ar
24	137	3.1	1156	16 Q66878	Q66878 aquifex aeo
25	136	3.1	1133	5 Q21022	Q21022 caenorhabdi
26	136	3.1	3419	11 Q55147	Q55147 rattus norv
27	135	3.1	4151	5 Q96936	Q96936 drosophila
28	135	3.1	5201	5 Q9U479	Q9u479 drosophila
29	135	3.1	5380	11 Q91ZU7	Q91zu7 mus musculu
30	135	3.1	5385	5 Q9V6V3	Q9v6v3 drosophila
31	135	3.1	5496	5 Q9V6V2	Q9v6v2 drosophila
32	135	3.1	7389	11 Q91ZU6	Q91zu6 mus musculu
33	135	3.1	8805	5 Q9V6V4	Q9v6v4 drosophila
34	134.5	3.1	1084	4 Q9NV11	Q9nv11 homo sapien
35	134.5	3.1	1362	4 Q9ULF8	Q9ulf8 homo sapien
36	134	3.1	1314	3 Q9C499	Q9c499 neurospora
37	133.5	3.1	775	4 Q96CN9	Q96cn9 homo sapien
38	133.5	3.1	2331	5 Q22847	Q22847 caenorhabdi
39	133.5	3.1	3263	5 Q917U3	Q917u3 drosophila
40	132.5	3.0	1171	3 Q9P3E2	Q9p3e2 neurospora
41	132.5	3.0	1627	5 Q96ZQ0	Q96zq0 giardia lam
42	132	3.0	4063	5 Q02425	Q02425 caenorhabdi
43	131.5	3.0	809	11 Q9D4H2	Q9d4h2 mus musculu
44	131.5	3.0	1313	10 Q9XIP6	Q9xip6 arabidopsis
45	131.5	3.0	1864	10 Q9FYL7	Q9fyl7 arabidopsis

ALIGNMENTS

RESULT 1
ID Q9LKN6 PRELIMINARY; PRT; 848 AA.
AC Q9LKN6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NUCLEAR CAP-BINDING PROTEIN CBP80 (mRNA CAP BINDING PROTEIN).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Kmiecik M., Jarmolowski A.;
RT "A nuclear cap-binding protein complex from Arabidopsis thaliana.
RT Molecular characterization of Arabidopsis thaliana CBP80.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Hugouvieux V., Kwak J.M., Schroeder J.I.;
RT "An mRNA Cap Binding Protein, ABH1, Modulates Early Absciscic Acid
RT Signal Transduction in Arabidopsis.";
RL Cell 106:477-487(2001).
DR EMBL; AF268377; AAK76167.1; -
DR EMBL; AF272891; AAK91588.1; -
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR003890; EIF4G_cent.
DR SMART; SM00543; MIF4G; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN.1.
SQ SEQUENCE 848 AA; 96547 MW; 1E583391F4BAE1A8 CRC64;

Query Match 100.0%; Score 4374; DB 10; Length 848;
Best Local Similarity 100.0%; Pred. No. 9.9e-293;

Matches 848: Conservative 0: Mismatches 0: Indels 0: Gaps 0:			
QY	1	MSNKTLLLRIGEGPEYGTSSDYKDHIETCFGVI	REIERSGQVLPFLQCAEQLPKH 60
DB	1	MSNKTLLLRIGEGPEYGTSSDYKDHIETCFGVI	REIERSGQVLPFLQCAEQLPKH 60
QY	61	IPLYGTILGLNLENEFVQKLVESVHANFOALD	SGNCNIRILLRFTMTSLCSKVQP 120
DB	61	IPLYGTILGLNLENEFVQKLVESVHANFOALD	SGNCNIRILLRFTMTSLCSKVQP 120
QY	121	ASLIVVFETLLSSAATTVDEKGNPSQPOADFV	ICILSLPWSGSELAEQVPDETERV 180
DB	121	ASLIVVFETLLSSAATTVDEKGNPSQPOADFV	ICILSLPWSGSELAEQVPDETERV 180
QY	181	LVGIQAYLSIRKNSSTSLGNFFHNGEFESSLA	EKDFVEDLLDRIQSLASNGWKLESVPRP 240
DB	181	LVGIQAYLSIRKNSSTSLGNFFHNGEFESSLA	EKDFVEDLLDRIQSLASNGWKLESVPRP 240
QY	241	HLSEFAQLVAGKFHELPIKCMEQSPSPSDH	SRAYSQKHDAITRYPQIRRLNIFPAN 300
DB	241	HLSEFAQLVAGKFHELPIKCMEQSPSPSDH	SRAYSQKHDAITRYPQIRRLNIFPAN 300
QY	301	KMEDVQPIDRFVVEEYLLDVLFLYNGCRKE	CASYMANLPVTFRYEYLMATLSQIILLP 360
DB	301	KMEDVQPIDRFVVEEYLLDVLFLYNGCRKE	CASYMANLPVTFRYEYLMATLSQIILLP 360
QY	361	QPPFKTYLTYLTVIMDLCKALPGAPPAVVA	GAVRALFEKISDLDMESRTRLIWFSSHLSN 420
DB	361	QPPFKTYLTYLTVIMDLCKALPGAPPAVVA	GAVRALFEKISDLDMESRTRLIWFSSHLSN 420
QY	421	FOIWPWEWAFVLDLPKWAPKRVFVQEILO	REVRLSYWDKIKOSIENATALEELLPPKA 480
DB	421	FOIWPWEWAFVLDLPKWAPKRVFVQEILO	REVRLSYWDKIKOSIENATALEELLPPKA 480
QY	481	GNPFMYSLSEEGKTEEQQLSAELSRKVK	KEQTARDIMVIEETIYPVHGFEVTLTIWQ 540
DB	481	GNPFMYSLSEEGKTEEQQLSAELSRKVK	KEQTARDIMVIEETIYPVHGFEVTLTIWQ 540
QY	541	TLLDIGSKSFTHLVTLERYGVQVFSKLC	PDNDKQVMLLSQVSTYWKNNVQMTAVIDRM 600
DB	541	TLLDIGSKSFTHLVTLERYGVQVFSKLC	PDNDKQVMLLSQVSTYWKNNVQMTAVIDRM 600
QY	601	GTVLSNQAIWRVWFSPENVDQFHVSDQP	WEILGNALNKTYNRISDLRKDISNITKNVLY 660
DB	601	GTVLSNQAIWRVWFSPENVDQFHVSDQP	WEILGNALNKTYNRISDLRKDISNITKNVLY 660
QY	661	AEKASANARVELEAAESKLSLVEGEPVL	GENPAKMKRLKSTVEKTEGAELSLSRESLEAKE 720
DB	661	AEKASANARVELEAAESKLSLVEGEPVL	GENPAKMKRLKSTVEKTEGAELSLSRESLEAKE 720
QY	721	ALLNRALSETEVLLLLFQSFGLVKERLP	DPPTKVRVSDQLKSTIGAEDDKPSAMDVDSEN 780
DB	721	ALLNRALSETEVLLLLFQSFGLVKERLP	DPPTKVRVSDQLKSTIGAEDDKPSAMDVDSEN 780
QY	781	GNPKKCEVGERQWCLSTLGYLTATROYA	SEIWPHEMKLESEVSGEDVHFLFQAIS 840
DB	781	GNPKKCEVGERQWCLSTLGYLTATROYA	SEIWPHEMKLESEVSGEDVHFLFQAIS 840
QY	841	SALQFPLH 848	
DB	841	SALQFPLH 848	
RESULT 2			
Q9SIU2			
ID	Q9SIU2	PRELIMINARY;	PRT; 749 AA.
AC	Q9SIU2;		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DE	POTATIVE CAP-BINDING PROTEIN.		
OS	AT2G13540.		
OS	Arabidopsis thaliana (Mouse-ear cross).		

OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxID=3702;
OX	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA;
RX	MEDLINE=20083487; PubMed=10617197;
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA	Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA	Salzberg S.L., Fraser C.M., Venter J.C.;
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT	thaliana.";
RL	Nature 402:761-768(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA;
RA	Lin X.;
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AC007063; AAD22677.1; -
DR	InterPro; IPR002106; AA_TRNA_ligase_II.
DR	InterPro; IPR003890; EIF4G_cent.
DR	Pfam; PF02854; MIF4G; 1.
DR	SMART; SM00543; MIF4G; 1.
DR	PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ	SEQUENCE 749 AA; 85610 MW; 8AF328F08CD33CB0 CRC64;
Query Match 86.5%; Score 3784; DB 10; Length 749;	
Best Local Similarity 99.1%; Pred. No. 3.6e-252;	
Matches 735; Conservative 2; Mismatches 5; Indels 0; Gaps	
Qy	1 MSNKTTLLLRIGEGPEYGTSSDYKDHIETCFGVIRREIERSGQVLPFLQCAEQLPKH 60
Db	1 MSNKTTLLLRIGEGPEYGTSSDYKDHIETCFGVIRREIERSGQVLPFLQCAEQLPKH 60
Qy	61 IPLYGTILGLNLENEFVQKLVESVHANFOALD SGNCNSTRILLRFTMTSLCSKVTP 120
Db	61 IPLYGTILGLNLENEFVQKLVESVHANFOALD SGNCNSTRILLRFTMTSLCSKVTP 120
Qy	121 ASLIVVFETLLSSAATTVDEKGNPSQPOADFVYICILSLPWGSELAEQVPDEIERV 180
Db	121 ASLIVVFETLLSSAATTVDEKGNPSQPOADFVYICILSLPWGSELAEQVPDEIERV 180
Qy	181 LVGIQAYLSIRKNSSTSGLNFFHNGFEFESSLA EKDFVEDLLDRIQSLASNGWKLESVPRP 240
Db	181 LVGIQAYLSIRKNSSTSGLNFFHNGFEFESSLA EKDFVEDLLDRIQSLASNGWKLESVPRP 240
Qy	241 HLSFEAQLVAGKFHELPIKCMEQSPSPSDHSRAYSGKQKHDAITRYPQIRRLNIFPAN 300
Db	241 HLSFEAQLVAGKFHELPIKCMEQSPSPSDHSRAYSGKQKHDAITRYPQIRRLNIFPAN 300
Qy	301 KMEDVQPIDRFVVEEYLLDVLFLYNGCRKECASYMANLPVTFRYFYLMAETLFSQILLP 360
Db	301 KMEDVQPIDRFVVEEYLLDVLFLYNGCRKECASYMANLPVTFRYFYLMAETLFSQILLP 360
Qy	361 QPPFKTYLTYLTVIMDLCKALPGAPPAVAVAGVRAALFEKISDLDMESRTRLILWFSSHLSN 420
Db	361 QPPFKTYLTYLTVIMDLCKALPGAPPAVAVAGVRAALFEKISDLDMESRTRLILWFSSHLSN 420
Qy	421 FQIWPWEWAFVLDLPKWAPKRVFVQEILOREVRLSYWDKIKOSIENATALEELLPPKA 480
Db	421 FQIWPWEWAFVLDLPKWAPKRVFVQEILOREVRLSYWDKIKOSIENATALEELLPPKA 480
Qy	481 GPNFMYSLEGGKTEEQQLSAELSRKVKKEQTARDIMVIEETIYPVHGFEVTLTIWVQ 540
Db	481 GPNFMYSLEGGKTEEQQLSAELSRKVKKEQTARDIMVIEETIYPVHGFEVTLTIWVQ 540
Qy	541 TLLDIGSKSFTHLVTLVLYRGVQVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVADRMM 600

```
Db 541 TLLDIGSKSTHLTVLRYGVESKLCPCNDKQVLLSQVSTYKNNVQMTAVADRMM 600
Qy 601 GYRLVSNQAIWRVSPENVDQFHVSDQWELIGNALNKTYNRISDLRDKISNITKNVLV 660
Db 601 GYRLVSNQAIWRVSPENVDQFHVSDQWELIGNALNKTYNRISDLRDKISNITKNVLV 660
Qy 661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRKLSTVEKTEGEAELSRESLEAKE 720
Db 661 AEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRKLSTVEKTEGEAELSRESLEAKE 720
Qy 721 ALLNRALSETEVLLQLLQFSL 742
Db 721 ALLNRALSETEVLYISLMQHMI 742

RESULT 3
Q9AWB3 PRELIMINARY; PRT; 910 AA.
AC Q9AWB3;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE NUCLEAR CAP-BINDING PROTEIN CBP80.
OS Oryza sativa subsp. japonica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Derba M., Kmiecik M., Jarmolowski A.;
RT "Molecular characterization of Oryza sativa CBP80.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY017415; AAG54079.1.
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR003890; EIF4G_cent.
DR InterPro; IPR02052; N6_Mtase.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
SQ SEQUENCE 910 AA; 103823 MW; 23D62F81E8596F56 CRC64;

Query Match 63.0%; Score 2753.5; DB 10; Length 910;
Best Local Similarity 61.3%; Pred. No. 4.7e-181;
Matches 524; Conservative 136; Mismatches 176; Indels 19; Gaps 5;

Qy 2 SNWKTLLLRIGKGPYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLQCAEQLPHKI 61
Db 3 AGWRTLLLRIGRCPEYGGVDHKEHETCYGLVCREYEHKDMAFEFLLQCADQLPHKI 62

Qy 62 PLYGTGLGLNLENDFFQKLVESVHANFQVALDSGNCNSIRILLRFMTSLSCSKVIOPA 121
Db 63 PFFGVGLIGLINEDFSKGIQVTDTHANLQDALHNENRDRILLRFLCGLCMCKSVLPN 122

Qy 122 SLIVVFETLLSAATVDEEKNQSPQADPVYTCILSSLPWGGSELAQVDEIERVL 181
Db 123 SIITFEALLSSAATILDEETGNPSQPRADPVYVCILASLPWGGSELFQVDFEERVL 182

Qy 182 VGIOAYLSIRKNSSTSGLNFFHNGEFESLSAEKDFVEDLLDRIQSLASNGWKLSEVPRPH 241
Db 183 VGIOYIISIRHFDIAFSVFTDE-GNSPNKKDFIEDLWRIQVLSRNGWVKVSPKPH 241

Qy 242 LSFEAQLVAGFHELRPIKCMQEQPPSDHSRAYSGKOKHDALTRYPORIRELNIFPANK 301
Db 242 LSFEAQLVAGVSHRSPISCPHLLSRNAS-SEIVGQKEADLRYPQRLRLHIFPTNK 300

Qy 302 MEDVQPIDRFVVEEYLLDVLVYNGCRKECASMANLPVTFRYEYLMATLFSQILLPQ 361
Db 301 AENMQPVDREVVEECILDVLLFFNCRKECAFVLSLPVFFRYEYLMATLFSQILLPN 360
```

```
Qy 362 PPEKTYLTVLIMDLCKALPCAPVAVGAVRALPEKISDLDMESRTRLILWFSSHLSNF 421
Db 361 PPERPYTYLTVLIMDLCKALPCAPSVVYGAVHALFDRISNMDRCRTRLILWFSSHLSNF 420
Qy 422 QFIWPEEWAFAVLDLPKAPKRVFOEILQREVRLSYWDKIKOSIENATALEELLPKAG 481
Db 421 QFIWPNQEWAYKDLPEKAPORVFOEVLREIRLSYFDKIKOSIEDAVELEELLPKAG 480
Qy 482 PNFMSLESGKTEKEEQSAELSRLKVEKQATARDMIVTIEETIYPVHGFEVTLTIYVOT 541
Db 481 PNFYHSDGKSTGDGHRLSKELVAMVRGRKTOGDIISWDEKIIPVNGAKFALDVVYOT 540
Qy 542 LLDIGSKSTHLTVLRYGVESKLCPCNDKQVLLSQVSTYKNNVQMTAVADRMM 601
Db 541 LLDIGSKSTHLTVLRYGVESKLCPCNDKQVLLSQVSTYKNNVQMTAVADRMM 600
Qy 602 YRLVSNQAIWRVSPENVDQFHVSDQWELIGNALNKTYNRISDLRDKISNITKNVLV 661
Db 601 YRLVSNQAIWRVSPENVDQFHVSDQWELIGNALNKTYNRISDLRDKISNITKNVLV 660
Qy 662 EKASANARVELEAAESKLSLVEGEPVLGENPAKMKRKLSTVEKTEGEAELSRESLEAKE 721
Db 661 KEASEKAARELEAKSIIIEVDGQVPSNPGRRLRQARADKAKEGEVTEESLEAKE 720
Qy 722 LLNRALSETEVLLQLLQFSLGVLKERLP-----DPTKVRVQDLKSGIAEDDKPSAMD 775
Db 721 LLARGLESKELLRLLFKSFVEYLTLPISADGVPNLRAGDPNPNVSSARDEATYME 780
Qy 776 VDSNG-----NPKKSCVGEREQWCLSTGLTGLTFTROYASEIHPHEKLESEV 825
Db 781 IDNENGDNDSQLNGQNKKISHNVGELBQWCLCTGLYKLSFSQYATEIWSHIAMLQOEI 840
Qy 826 FSGEDVHPLFLQAIS 840
Db 841 FVG-NIHLIRKAAS 854

RESULT 4
Q9U980 PRELIMINARY; PRT; 800 AA.
AC Q9U980;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CAP BINDING PROTEIN 80.
GN CBP80 OR EG:84H4.3 OR CG7035.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=OREGON R; TISSUE=IMAGINAL DISC;
RA Lewis J.D.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238970; CAB53186.1.
DR FlyBase; FBgn002942; Cbp80.
DR InterPro; IPR003890; EIF4G_cent.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
SQ SEQUENCE 800 AA; 93070 MW; 7C9E7948ABD876B6 CRC64;
```

```
Query Match 18.7%; Score 818.5; DB 5; Length 800;
Best Local Similarity 25.6%; Pred. No. 7.7e-48;
Matches 218; Conservative 175; Mismatches 342; Indels 115; Gaps 19;

Qy 5 KTLRLRIGKGPYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLQCAEQLPHKIPLY 64
Db 35 ESLILRVGE-----ATTSSVESNLEGLVSLVLEADLTGTRFLKILRLISDCAVAMPEKTVY 89
Qy 65 GTLIGLLNLENDFFQKLVESVHANFQVALDSGNCNSIRILLRFMTSLSCSKVIOPA 124
```



```
Db 35 ESLILRVGER-----STSVESNLEGLSVLEADLGTFRKLILRLSDCAVRMPKCTVY 89
Qy 65 GTILGLLENEDFDYOKLVESVHANFOVALDSGNCNIRILLRFTWTLSCSVIOPASLI 124
Db 90 TTLVGLLNKAKYKFGGEFVDHVMVTKESLKMCRDAARYSLRFLADLVNCHVISATSLL 149
Qy 125 VVFETLLSAATVDEEKGNSQWQADPFYVICILSSLPWGSSELAEOVPDPIERVVLGI 184
Db 150 QLLDTPID-----VSNEDTVP--QVRDWFVFAVLSTLPWVGROLYEKESALESLLRI 202
Qy 165 QAYLSIRKNSSTGLNFHNGEFESLAEKDEVEDLLRIQSLANGKWLKSVPRPHLSF 244
Db 203 EYLKNSKKHNNALRVSS-----DAPHQBEYLDCLMAQIRKLQDNNAEKHPRPYLVE 259
Qy 245 EAQVAGFHLELRPTKCMEOQSPSDHSGRAYSGKQKHDALTRYQRIIRLNIFFPANKMED 304
Db 260 DSILCEALQHLPTI-----VPPPHDN-----FEYPMWVYVYRFDYDTPD 302
Qy 305 -----VOPIDRFVVEYLDDVLFYNGCRKECASYMANLPVTFR--YEYLMATLFSOI 356
Db 303 GPNLPGASHIERFLIEEHLHIIETIYHHERKDCAQLLSFPYKHKIPLYEIVVFAEL 362
Qy 357 LLLPQPPFTLYTIVIMDLKALPGAPVAVAGAVRALFEKISDLDMESRTRLINFSH 416
Db 363 FHMTPTRYLDICYGSIILIECLKLQAPLTPQVLAQAATEILFMRIDSMNTSCDFRVNWFYS 422
Qy 417 HLSNFOFTWPEEAWFVLDLPKPAKRVFVQEILOREVRLSYMDKIKOSIENATALEELL 476
Db 423 HLSNFKFTWDEWDSCLLDGERPRPFIOEVLQCKRLSYHQRTMM--PTTYAKLI 480
Qy 477 PPKAGPNMYSLIEGKEKTEEQLSAELSRVKYKQRTARDMIVWIEETIYPVHGFE--- 532
Db 481 PLTPVPNYKANEAA--ANLPGTTVAHOLVVAIROKCTPEEVNVLKD--IPNSGYSGEEM 537
Qy 533 -----VTLIVVOTLIDIGSKSTHLVTLVRLYGVQVSKLCPDNDQVMLLSQVSYW 585
Db 538 SDGSFNALKIDVFQTLNLGSKSPSFSAATSKFHSFALAEAEAEQICILHNFELW 597
Qy 586 KNNVQMTAVATDRMGYRLVSNQATVRVVFSPENVQDFHVSQDPEWELLGNALNKYNRIS 645
Db 598 SSHOQMVVLIDKLKQIVCSAVATWIFSKEMTGEF--TKLYLWEILHLTIKKRKHVI 656
Qy 646 DLKRDINITKNVLVAEKASANARVELEAESAELSLVEGEVPLGPNKMRKLKSTVEKT 705
Db 657 KLNTELSEAKELAKADSSDS-----EDDSSHRKKKPIIT--- 692
Qy 706 GEALUSRESLEAKALLNRLALSETEVILLLLFQSLGVLKERLPDPTKVRSVQDLKSG 765
Db 693 -HADKPSSEVVERMEKLEAANVOKRLFLIVQRFIMILSEHL-----LRSDDT----- 741
Qy 766 AEDDKPSAMVDSENGNPKKCEVGERQWCLSTLGYLTAFTRQYASEIWPHEKLESEV 825
Db 742 -----GRDPDTPD-----WYRWTTIGRLQOVFLMHHEQVQKYSSTLETIL 779
Qy 826 FSGE-DVHPL 834
Db 780 FTSDDLTHIL 789

RESULT 6
O01763
ID O01763 PRELIMINARY; PRT; 754 AA.
AC O01763
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 87.7 KDA PROTEIN.
GN F37E3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RL investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wamsley P., Bradshaw H.;
RT "The sequence of C. elegans cosmid F37E3."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003132; AAB54133.2; .
DR InterPro; IPR000179; Cyt.b.b6.
DR InterPro; IPR003890; EIF4G-cent.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 754 AA; 87715 MW; FFBA943CBE23876 CRC64;
```

Query Match 14.6%; Score 636.5; DB 5; Length 754;
Best Local Similarity 24.4%; Pred. No. 2.4e-35;
Matches 182; Conservative 163; Mismatches 322; Indels 79; Gaps 20;

```
Qy 11 ICEGPEYTSDDYKDIETCFGVIRRETERSDQVLPFLQCAEQLPKIPLYGTGL 70
Db 35 IKGVGENTGSSIEC--NLDKLTAFLHDDLEKVRASIIDIAAGCAIYLPNRTVYTLVL 92
Qy 71 LNLNEEDFVKLVESVHANFOVALDSGNCNIRILLRFTWTLSCSVIOPASLIVPETL 130
Db 93 LNSKNFNGDGVVEKLISQQLLSKQYOEAOQLAIFLCLDNGSVGLTAQSIGEYLESF 152
Qy 131 LSSAATTVDEEKGNSQWQADPFYVICILSSLPWGSSELAEOVPDPIERVVLGQAYLSI 190
Db 153 IAAAF-----EENMP--QVRNDYIOTVRLCLPWICKELTEKAPEQMEIGALGYLEL 205
Qy 191 RKNSTSGUNFPHNGEFESLAEKDEVEDLLRIQSLANGKWLKSVPRPHLSFEAQLVA 250
Db 206 RKNHVALQVWREGSTQK--QEDYLESLSAQIEALRNADVMVENHIPRHSYGFETTLQD 263
Qy 251 GKFHELRLPTKCMEOQSPSDHSGRAYSGKQKHDALTRYQRIIRL-----NIFPANKMED 304
Db 264 ALQHLNLPSPQSPEHTS-----DMYYPYLVVFLFQDQADCSAESSKPLPG 308
Qy 305 VOPIDRFVVEYLDDVLFYNGCRKECASYMANLPVTFR--YEYLMATLFSOI 357
Db 309 DSSIDRFLEGEIAWIEKQFNRRKACARELLAFAPENSPVIGF-----LIFETIQOML 364
Qy 358 LLPQPPFTLYTIVIMDLKALPGAPVAVAGAVRALFEKISDLDMESRTRLINFSH 417
Db 365 RLPHAPYPAIFHCSLVLELLKLPDDYPOILVQTVCEIYRADSMQVPCIDRMVDWFSFH 424
Qy 418 LSNFOFTWPEEAWFVLDLPKPAKRVFVQEILOREVRLSYMDKIKOSIENATALELLP 477
Db 425 LSNFOYRYTWTQDKLCKNDAFSGSQIFVREVIEKCRREGSKYKIAALPQ--DFVKIHP 482
Qy 478 PRAGPNMYSLIEGKEKTEEQLSAELSRVKYKQRTARDMIVWIEETIYPVHGFE 531
Db 483 --CSPEVRYLDE--EDTALVQRAETFTTFQFQEPAPAEFLNELKSNDENDLPPYINEF 538
Qy 532 EYTLIVVOTLIDIGSKSTHLVTLVRLYGVQVSKLCPDND--QVMLLSQVSYWKNVQ 590
Db 539 ---CLFVVMVLMKASKYSHNFSALFRQTTLTKTVCDASELYOEKLETLISCKWNTNQ 594
```

QY 591 MTAIVADRMGMRYLVSNOAIVRWVSPENVDQFHVSDPW--EILGNALNKTNYNRISDLR 648
 Db 595 MLMLITDKLLKMQVIDCSAVVWGLFD-EKMQEH--DQWLFVNLQALEKLTRQINVE 651
 QY 649 KDISNITKNVLAEKASANARVELEAAEKSLSLVEGEPVLGNPAKMKRLKSTVEKTEA 708
 Db 652 KDKEKTEK-----TENIKKEDEDESKIMDEDE-----TKKEKFKQDLELDENN 697
 QY 709 ELSIRLESLEAKALLNRLSETVLL 734
 Db 698 KEXLERWVTKQGLFNDFLHAETLL 723

RESULT 7
 Q9VDA4 PRELIMINARY; PRT; 747 AA.
 AC Q9VDA4;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE CG7907 PROTEIN.
 GN CG7907
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolahkov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003734; AAF55893.1;
 DR FlyBase: FBgn0038887; CG7907.
 DR InterPro: IPR003890; EIF4G_cent.
 DR Pfam: PF02854; MIF4G; 1.

DR SMART: SM00543; MIF4G; 1.
 SQ SEQUENCE 747 AA; 85089 MW; 5FC51461A12057E6 CRC64;

Query Match 12.6%; Score 550.5; DB 5; Length 747;
 Best Local Similarity 25.0%; Pred. No. 2e-29;
 Matches 167; Conservative 136; Mismatches 257; Indels 107; Gaps 22;

QY 7 LLIRIGEGPEY---GTSSDYKDHIETCFGVRIRTERSQDVLFFLQCAEQLPKHIP 62
 Db 45 LELKGNIG--YLLGKGSVDLRENL-----INELVICLGYYPGQAS 84
 QY 63 LYGTLLGLNLENEDE-----VQKLVSVEHANFOALDSGNCNSIRILLRFTSLIC 114
 Db 85 AYATLVGLINVADEFGEGECCLLFNAHKVGESVHIR-----DNNRCGVVHCNVLYH 136
 QY 115 SKVIQASLIVVFETLLSSAATTVD--BEKGNPSWQPADFYVICILSLPGLGGSELARQV 173
 Db 137 CQVLPSVYILKLLAAFLKDCALKDDDLGVTPQIRRDFLAYCVLSAMPLIGRDEGET 196
 QY 174 PDEIERVLGTOAYLSIRKNSSTSGLNFFHNGEFESSLAEEKDFVEDLLDRIOSLASNGW- 232
 Db 197 --AFDKLIVSLQIYIKKRSALHTNLSVW-----PDFNORDYLELLMQQVDGMRQHW 248
 QY 233 --KLESVPRPHLFEAOLVAGKPHLRPIKMEQSPSPSDHSRAYSGKQKHDALTRYPOR 290
 Db 249 EPEHQILPRPKSFSETLSHGRHQLR-----DYDLA-----AHEERCYPLP 291
 QY 291 IRRNLTFPANKMEDVQ-----PIDRFVVEYLLDVLFLNGCRKECAS-----YMANLPV 340
 Db 292 RVCFRIFSCDSVGEIPNMPVPVSTIERHLEAHILDLISFHKERKICADSLMLVAASKPO 351
 QY 341 TFRYEYLMATLSQILLPQPPTLYTIVIMDLCKALPGAPVAVGAVRALFEKIS 400
 Db 352 LPVI-YCIVEVILGEMLRPLTANWSTIAYGSLVELCKRQPKIPQVVAQADIIYNRLN 410
 QY 401 DLDMESTRLILWFSHLSNFQFIWPEWEAFVLDLP-----KWAPKRVFVQETILOREV 454
 Db 411 SMSVACFDRLVNVVSHHSINFGFCNQSKWAQGLSPIDPSATNLQPKVVELRELLKKCF 470
 QY 455 RLSYWDKIKOSIENATALELLPPKAGPNFMYSEEGKEKTEEQOLSSELKRYKEKOTA 514
 Db 471 R-----IKDVPD--TLADFLPPPLPHKFEV---DETLPGAILSKDLEAMRSPQAS 518
 QY 515 RDMVTWEETIYPVHGFEVTLTIIVVQTLIDIGSKSFTHLVTLVRLERYGOVFSKL-CPDNDK 573
 Db 519 PEMISEIKSTGI-GPLLKINVTQNLHLGSKSFSTHTFGILRKYHSVFDLVAGDPER 577
 QY 574 QVMLLSQVSTYWKNNVQMTAVADMGMGYRLVSNOAIVRWVFPSP--ENVDFHVSQDPW 630
 Db 578 QAAVLNGIFDWWASDQYKFEVTEKLVTLGLIEPISVVRWIFGFSMRKELTKIYI---W 633
 QY 631 EILGNAL 637
 Db 634 ELLHSAL 640

RESULT 8
 Q9HE91 PRELIMINARY; PRT; 829 AA.
 AC Q9HE91;
 DT 01-MAR-2001 (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE RELATED TO CAP BINDING PROTEIN 80 (CBP80).
 GN B13020.110.
 OC Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.,
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL451015; CAC18201.1;
SQ SEQUENCE 829 AA; 92651 MW; BFBABCBB76304ACB CRC64;

Query Match 11.5%; Score 501.5; DB 3; Length 829;
Best Local Similarity 21.6%; Pred. No. 5.6e-26;
Matches 176; Conservative 156; Mismatches 319; Indels 165; Gaps 26;

QY 50 LLOCAEQPHKIPLYGTGLGNLENEFVQKLVSVHANFQVALDSCNCSIRILLRFM 109
DB 86 VLQAVEOPMTFPLAAVVLVANTKNKPOIVDMLLAKLTSTETKTAEGDWRDVKLYLKA 145
QY 110 TSLCSKVQIASLIWVFETLLS-----SAATTVDEEKGNSWQPADFYVICILSS 161
DB 146 GCL--QSCLEGDGVFVPLEELFSRAVDLQTASSEDTIGTE-----LVKILLT 191
QY 162 LPW-----GGSELAEOVPDEIR--VLVG-----IQAYLSIRKNSSTGLNFHNGFESSL 211
DB 192 LPYVMVTGPTQWQQAADLMETKEITIAEHAHPLQALIE-----PYHPEAGDQSPV 241
QY 212 AEKDFVEDLLDRIOQLASNGKLESVPRP-----HLSFEAQLVAGKEHLELRPIKME-- 263
DB 242 ASQSCISLLOLQONEANGWALSCLPRPWEVVEEQEQAANAERHTLPPISIPQTV 301
QY 264 ---QSPSPSDHSRAYSQKQKHDALTYPQRIIRLNIFPANKMEDVQPIDRF--VVEEY 316
DB 302 VAGRPRLPEIYFSYANQD-----VSTPTPLNTIAASLIIDA 339
QY 317 LLDVLFYNGCKEACASYMANL-----EVTF-----RYEYLMAET 351
DB 340 LVDITNVLNKNNVTARELIELDCYFAPRTFAARATPDKIREIPPXSKMKPEDVAVDA 399
QY 352 LFSQILLLPQPFKLYTLVIMDLCKALPGAFPAVAVAGVRALEPKISDLDMSERTRI 411
DB 400 VFSGLFSLPNEHKLIVYHAYLVEACKLAPAAIAFSLGRAIRFLYRNHRMDLELSNRFM 459
QY 412 LWFSHLSNFOFIPWEEWAFVLDLPKWKPRVFQETILOREVRLSYWDKIKQSTENATA 471
DB 460 DWFSHLSNFOFTHKWTEDVNVLSNHPDKAFINGAIDREIRLSFAQRIKNTLPE--P 517
QY 472 LEELLPP---KAGPNFYSLSEEGKEKTEEQQLSAELSRKVKKEKQATQADMIVMIEETIYPV 528
DB 518 YQSLIGPEKEKDVDPKFAADKTPPAEGKETAALLRRKAPEE-----IEPVIERI 569
QY 529 HGFEV-----TLTIVQTLIDIGSKSTHLYTVLRYGVQVFSKLCPPND--KQVMLL 578
DB 570 HSLADNNLDPLVASTDVFVSVLHGSKSLSHVLAARTKERTLDAGATSEAAQTQII 629
QY 579 SQVSYKNNYQMTAVADRMGMGRVLSNQAIKRVWVSPENVDPQHVSDQWPEILGNALN 638
DB 630 SSVMEYSAHPGVAIAIEKLLNYSILTPQAVINWA-----ITYAGATGEALA 679
QY 639 KTYNRISDLRKDISNITKNLVIAEKASANARVELEPAESKISLVEGEVPLGENPAKMKRL 698
DB 680 KGF-----VYEMVNTVV--KVTSLRQLQKATLPEAMIDDETEAEING--MRSL 727
QY 699 KSTVEKT--GAEULSRLSEAKAL--LNRLSETEVLLLLLQSFGLVKER----- 748
DB 728 FRAIEDALFAWASGSKDMEASDGLGEGDGTSETKLVKRWGRWLVRFRRAAIEEAF 787
QY 749 LPDPTKVR-----SVODLKSTIGADDKPSAMVDVS 778
DB 788 IVEARKERKRAAAAALAAAPAGEDDMVAVDSVDA 823

RESULT 9
O14253

ID O14253 PRELIMINARY; PRT: 780 AA.
AC O14253;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL 90.1 KDA PROTEIN C6G10.07 IN CHROMOSOME I.
GN SPAC6G10.07.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST GCR3.
DR EMBL; Z98603; CAB11293.1;
DR InterPro: IPR003890; Eif4G_cent.
DR InterPro: IPR002173; PfkB.
DR Pfam: PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
DR PROSITE; PS00584; PEKB_KINASES_2; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 780 AA; 90077 MW; 6AAA6562287F74F7 CRC64;

Query Match 10.9%; Score 475.5; DB 3; Length 780;
Best Local Similarity 23.9%; Pred. No. 3.2e-24;
Matches 192; Conservative 149; Mismatches 299; Indels 163; Gaps 36;

QY 40 ERSGQVLPFLQCAEQ-----LPHKIPLYGTGLGNLENEFVQKLVSVHANFOALD 95
DB 67 EANDTEITSLDICTQTTAFIIPVKIPLATLIIRASLR-----VPLILEKAAAYFCLOYF 122
QY 96 SGNCSNI-----RILLRFMTSLC-SKVTPASLIWVFETLLSSAATTVDEEKGNSWQ 149
DB 123 T-NLMSFLYKAVDLRM--LICMSFALQGTL-----KPLFSLADAKSKETKPSVW-- 172
QY 150 QADFVVICILSLPW-----GGSELAEOVPDEIRVLVGIQAYLSIRKNSST--SGL 199
DB 173 -GDNFRILINLPYFAANNDLGKKDFANEILDQCE-----IVVRHKSSITLSNPL 224
QY 200 NEFFHNGFESSLAEDKDFEDLLDRIOQLA-SNGWKLESVPRPHLSPEAQLVAGKEHLELRP 258
DB 225 SIHDN-----LSEEL--DLLYKQILSRENDFTFPYISQPMKFFESDFV----HIV-- 270
QY 259 IKCMGQPPSPSDHSRAYSQKQKHDALTYPQRIIRLNIFPANKMEDVQPIDRFVVEYLL 318
DB 271 -----PVSPSPTEWTFQTPQONELPSEKRFELFNFEIRTPDASDVAAISIFRDISV 324
QY 319 DVLFLYNGCKEACASYMANLPVTFY-----EYLMATLTF 353
DB 325 DVINHLFEENRVEAAQVLTLDVYFYTKTFALRTPVNLPLNLPDSESRNKAEDIIIEAVL 384
QY 354 SOILLPQPFKLYTLVIMDLCKALPGAFPAVAVAGVRALEPKISDLDMSERTRILLW 413
DB 385 GELLGQNTYKPVVYHSLLECCRIAPKILAPTFGRVIRLMYTMSSDLPLOQLDFIDW 444
QY 414 FSHLSNFOFIPWEEWAFVLDLPKWKPRVFQETILOREVRLSYWDKIKQSTENATALE 473
DB 445 FSHLSNFNHFKWNEIPDVELDDLHPKKVFNRETITRELLISYVTRISDLP-----E 499
QY 474 ELL----PPKAGPNFYSLSEEGKEKTEQQL--SAELSRKVKKEKQATQADMIVMIEETIYP 527
DB 500 ELRCLLGEGSPGNEVYENETHPLYQSSQIIEALRLHKPLEE---LDIILQSEE---- 551
QY 528 VHGFVET--LTIIVQTLIDIGSKSTHLYTVLRYGVQV----FSKLCPDNKKQVWLLSQVST 583
DB 552 IQNSETSAVRLVMSCAYSILGSRFSFHALNVEFKHLTLKHFSS--KSLSEIEVVDLEFS 609
QY 584 YHKNVQMTAVADRMGMGRVLSNQAIKRVWVSPENVQDFHVSQDPWEILGNALNKTYNR 643


```

Db 610 FWKLOPFNAVMWLDKMLNYSIIISITSIWLIK-QDVTIWSRS-YTWSL-----VNTTENK 663
QY 644 I-SDLRKDISNITKNVLAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTV 702
Db 664 LAARLRSSVKNEDSSLINE-----ANEKEIVTNLL-----LSALRALI 703
QY 703 EKTGEAELSRESLEAKAALLNRALSETEVLLLLFQS-FLGYLKERLPDPPTKVRVSQDL 761
Db 704 SENAE-----NIWVSHWLNLMKYVESNLSVKKOTIEANE--PVQEN 745
QY 762 KSIGAEDDRPSAMD-VD-----SEN 780
Db 746 TSEEDQETKQPVDVDAVDEQPSN 768

RESULT 10
QY717 PRELIMINARY; PRT; 886 AA.
AC QY717;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CACR3 PROTEIN.
GN CACR3.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO1060;
RX MEDLINE=99221032; PubMed=10206191;
RA Uemura H., Nakamoto K., Sugloka S., Tadenuma M.;
RT "Isolation and sequence of the GCR3 homologue from Candida albicans by
RT complementation of delta gcr3 strain of Saccharomycetes cerevisiae.";
RL Yeast 15:323-327(1999).
DR EMBL; AB015281; BAA76378.1; -.
DR InterPro; IPR003890; EIF4G_cent.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
SQ SEQUENCE 886 AA; 103127 MW; BEB223BDC864B4E CRC64;

```

Query Match 8.6%; Score 374.5; DB 3; Length 886;
 Best Local Similarity 21.5%; Pred. No. 3.5e-17;
 Matches 191; Conservative 143; Mismatches 313; Indels 241; Gaps 35;

```

QY 10 RIGEGPEYGTSDYKDIHETCFGVIIRREIERSGDQVLPFLQCAQPLPKIPLYGLIG 69
Db 51 RLGEAGDIENLIAD-TNYSINPIVAEFKIDNLRNLSILSTIYALITEQPHKISAIALIL 109
QY 70 LNLNEDFVQLVESVHANFQVALDSGN---CNSRILLRMTSLCSKVQPASLVV 126
Db 110 ICNAKNFVIKVIYELHSMOTMLDSQBSGVFNKILKFLSTL--TPIIEDNGIIQI 167
QY 127 FETLLSSAATTVDDEKGNPSQPOADFVYICILSSLPWGSSELAEQVPDEIER--VLVG 183
Db 168 FKQFL-NFAIEIQEQTEVRGLAQEIYNVLI--ALPY---VLSDNSDDLKINSINELIE 221
QY 184 IQAYLSIRKNSSTSGLNF-FHNGEFSSLAEKDFVEDLLDRIOSLASNGWK----- 233
Db 222 LARNEKIVAESATVLLPFDTRNGNGLPYIPKQWVDLILPALIKLQNDNPFRLFDFKT 281
QY 234 -LESVPRPHL---SFAQLVAGKFHELRIKMEQSPSPDSHRSYSGK----- 278
Db 282 HLDPVQSALENNSSISSELVKHKLPOL-----SFPVSNAEFSKTSIDKLWMDNPR 332
QY 279 ---QKHDALTRYPQIRRLNIPANKMEDVQPIDREV---VEEYLLDVLFLYNGRKECA 332
Db 333 YLFQVYNTTET-----ETVPPIETYLGLFFKFDIAFDILTILNLSFNKNETA 377
QY 333 SYMANLPV-----TFRVEYLMATFLFSQILL 358

```

```

Db 378 IELSILDMFEDNKLFAPPGTSIDOLNAYIEDNKSGETNDPSLSTWKIEDVAVESILTMIQF 437
QY 359 LPOPPKTYLTYYTIVLDLCKALPCAFPAVAGAVRALFEKISDLDMESRTRLLIWFSHHL 418
Db 438 LPNLEVEIYYIYTVLISCCSPESIAPVFGRAIRYFYNHLETJDYELKIRFLDMMSTQL 497
QY 419 SNFOFTWPEWAF-----VLDPKAPKRVFOBILOREVRLSYWDKIKQSI-----ENA 469
Db 498 SNFEKWKWEWSDSKKLKDL-KYHPKKNFIKNLIAKEIRLSNKKRKDKSFIDVIDGEV 556
QY 470 TALEE-----LLPPKAGPNFMYSLKEGKTEQ--OLSABELSKVKKQKOT--ARDM 517
Db 557 VNLDEFQYLDISMVPDVKSYIISYDTELYGESSRDLTIQIYEQQLNSKSNIGAGNE 616
QY 518 IVW-----IEETIYPVHGFEVT-----LTIVVQT 541
Db 617 IFFNFTNSELPPFHETASKVYDFILTHWKSNTDFNELYKSVLESTAPNNERFAINLIIT 676
QY 542 LLDIGSKSFTHLVTLERYGVFSKLCPPDNDKQVMLLSQVSTY-----WK 586
Db 677 YAIIGRSIYSVVSILSR-----DINKLKFLSGAPIDYVGDARFEDLHFTBEK 726
QY 587 NNQMTAVAIIDRMGYRLVSNQAIVR-WVFSP-----ENVQDFHVSQDPWEILGNALN 638
Db 727 QNRQNMII-----EAVFRIWHQPQVVFILVLEVLIEFGIID-PKYILVKALE 772
QY 639 KTYNRISDLRKDISNITKNVLAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRL 698
Db 773 S---NLIID-----NVSCMESIN---RI-LSKRAESKELIIQ-----LF 803
QY 699 KSTVEKTGEAELSRESLEAKEALLNRALSETEVLLLLFQSFGLVLK 746
Db 804 TAIIDNKLKLELDERKVEIEETESNA---TEVDKQWLFEYELGLLK 848

```

RESULT 11

```

QYXWR0 PRELIMINARY; PRT; 1413 AA.
AC QYXWR0;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Y11D7A.14 PROTEIN.
GN Y11D7A.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Steward C.A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL0332632; CAA21588.1; -.
DR HSP; P08799; LMND.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 1413 AA; 163882 MW; 0ED5080C8A6E5B6A CRC64;

```

Query Match 3.7%; Score 163; DB 5; Length 1413;
 Best Local Similarity 18.1%; Pred. No. 0.026;

Qy	338	LPV---TFRVEYLMAEYLFSGILLPPQPPKTYLYTLVIMDLCKALPAPVAVGAVRA	399
Db	451	MDEVQDTFRDEVKAKTSLOK-----DLEKATKNC--RILSKLKK	489
Qy	395	LFEKISOLDMESRTRLTLWFSSHLSNFOFTWPEWNAFVLDLPKWAPKRVFQEIILOREV	454
Db	490	SDRKIETLEORSS-----FNAELSN-----KIKKLEE	518
Qy	455	RLSYWDYIKOSIENATALEBLLPP--KAGPNF-----MYSLEGKEKTEQ	498
Db	519	ELRFSNELTRKIQ--AEAEEELNPGKKKAPMLGVLGKSTSADAKFTRESITRGGSQEDPQ	576
Qy	499	OLSAELSRKVKKEQTADMTWIEETIYPVHG-----FEVTLTIY-----	538
Db	577	HLQRELODSIERETDLKQKFAEEELQRLURDRERKRVFSCGTQTEVPLEVYAFPRGTQ	636
Qy	539	VOTLLDIGSKSTHLVT-----VLERYGQ	562
Db	637	TVATVQSDMSTVENLVTSNVAVTQDTDFEVPDRNVSIERETMSSPAGLFPSSSSRVGQ	696
Qy	563	VFKKLCPDNDQVMLLSQVSTYKNNVQMTAVADRMGMGYRLVSNQAIWRWF-----SPE	618
Db	697	SGRKLSPTPHRL-----APEVIADRDEGISDEDDPAELRILLELNEQEA	742
Qy	619	NVDQFHVSDQFWEILGNALNKTYNR--ISDLRKDINITKNVILVAEKASANARVE-----	671
Db	743	SILRLKVEDLEKE---NAESKVVYRELQAKLRQDSSNGSKSLSLGTSSAAEKVKVTL	799
Qy	672	-----LEAAESKLSIVEGE--PVLGPNPAKMKRLKSTVEKTKGEAEL	710
Db	800	NEELVOLURRTLTEKEQTVDSLKNQSLKDLTLETENDKLAKENKRLLALRKASEKTGEVDQ	859
Qy	711	SLRES-----LEAKEAL-----LNRLS	728
Db	860	KMESLAQAQERDELTA RLKRQLEAEDLPPTAKRVNDLTPKSHLKKWVELEDEIS	919
Qy	729	ETEVLL-----LLAPQSGFLVKERLPDPTKVR-----VODLKSIGAEDDKPSAMDV	776
Db	920	EMRVMLSSSGTQDLKALQSAKGALEEDLRCKQKLSLAEGDVQRLKLLNGSSSKVSELEQ	979
Qy	777	DSENGNPK-----KSCVEGREGQWCLSTLGLYATFTROYASETPHWEKLESEVFS	827
Db	980	KLKRGDEEAKKLSKLDKEDVKYKQEAQULG-----ETSKSTWESQSREKEKLS	1031
RESULT 13			
Q92B64	PRELIMINARY: PRT; 1023 AA.		
AC	Q92B64		
ID	Q92B64;		
DT	01-DEC-2001 (TReMBLrel. 19, Created)		
DT	01-DEC-2001 (TReMBLrel. 19, Last sequence update)		
DE	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	SRCC PROTEIN		
GN	SRCC OR LIn1686.		
OS	Listeria innocua.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Listeria.		
NCBI	Taxid=1642;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	STRAIN=CLIP 11262 / SEROVAR 6A;		
RX	PubMed=11679669;		
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,		
RA	Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,		
RA	Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,		
RA	Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,		
RA	Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,		
RA	Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,		
RA	Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkut G.,		
RA	Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,		
RA	Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,		
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tlierrez A.,		
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;		


```
QY 359 LPQPPKTYLTYLVIMDLKALPGAPVAVGAVRALFEKISDLDMESRTRLILWF---S 415
  || : : : || : : :
Db 1765 LDQ-----PGEVHHLKSSIRKLKVHI-DADEKKHQNIQLKESK 1804
  || : : : || : : :
QY 416 HH-----LSNQFIWPEWEAFVLDLPKAPRV----- 444
  || : : : || : : :
Db 1805 HHADLLKRVENLEQELIUSEKNMIFQAEKSKAEIOTLASEIORMAQNLODLOLELISTR 1864
  || : : : || : : :
QY 445 -----FVQRIQREVRLSYWDKIKQSIEN-----ATALEELL 476
  || : : : || : : :
Db 1865 SENENLIKELKEQOERVSDELTINPSIENLLDKDQEKVQMKREEAKITVEMLOTLQKEL- 1923
  || : : : || : : :
QY 477 PKAGPNFMYSLKEGK--KTEBQOLASLSRKVKK-KTARDMIYMIETIYPVHGFEV 533
  || : : : || : : :
Db 1924 -----NETVVSICNDQEVSKTRQNLGSOVOTLEKAQLLODLGEAKNKYII----FQS 1974
  || : : : || : : :
QY 534 TLTIVVOTL-----LDIGSKSFTHLVTLERYGVFSKLCPDNDKQ----- 574
  || : : : || : : :
Db 1975 SVNALTOEVEAGKQKLEKEKEIRTLKEQLKSOEQLVCKLAQVEGEQELWQOKLELRNV 2034
  || : : : || : : :
QY 575 VMLLSQ-----VSTYWKNNVQMTAVAIIDRMGMYRLVSNQAIYRWVFSPEVNDQ 622
  || : : : || : : :
Db 2035 THALEQKVQVLOSENNTLOSTY--EALONSHKSLESELGLIKLEKVALVERVST----- 2086
  || : : : || : : :
QY 623 FHVSDQPEI-----LGNALNKTYNRISDLRKDISNITKNVLVAEKASANARV----- 670
  || : : : || : : :
Db 2087 --ISGKEAELORELDMLQKTTQLSQEDYNKRNLTVEEVLEELQNTKAAHLKSVNQL 2144
  || : : : || : : :
QY 671 --ELEAAESKLSL-----VEGE-----PVLCENPA 693
  || : : : || : : :
Db 2145 EXELORAQOKIKMLKSCROLEGEKEMLQKELSQLAAQQOQAGSLVDSNVDEVMTEN-- 2202
  || : : : || : : :
QY 694 KMKRLKSTV-EKTGEAE-----LSLRESLEAKEALLNRLALSETEVLLILLFQS-- 740
  || : : : || : : :
Db 2203 --KALKETLEEKVKADKYLDKYCSLLISHHEELEKAKEIL-----EIEVARLKSRQSRQ 2254
  || : : : || : : :
QY 741 ---FLGVLKERLPDPTKVRSVODLKSIGABDDKPSAMDYDSENGNPKKSCYVGEREQWCL 797
  || : : : || : : :
Db 2255 DLQSSPLLNSSIPGSPNTPSVSEMK-----SASQONKASGK----- 2290
  || : : : || : : :
QY 798 STLGYLTATFROYASEIWPHEKML-----ESEVFS 827
  || : : : || : : :
Db 2291 -----RORSGVWEHGRAPSTAEFTS 2313
  || : : : || : : :
```

Search completed: August 20, 2002, 15:35:36
Job time: 579 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 15:26:32 ; Search time 36.79 Seconds
(without alignments)
892.476 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLRLRICKGPEYGT.....EDVHPLFLQAISALQFPPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	813	18.6	790	1 CB80_HUMAN	Q09161 homo sapien
2	279	6.4	861	1 GCR3_YEAST	P34160 saccharomyc
3	166	3.8	879	1 RA50_PYRHO	O58687 pyrococcus
4	140.5	3.2	539	1 MYS3_HYDAT	P39922 hydra atten
5	134.5	3.1	886	1 RA50_ARCFU	O29230 archaeoglob
6	134.5	3.1	1818	1 HMW2_MYCPN	P75471 mycoplasma
7	133.5	3.1	2363	1 SPCO_MOUSE	Q62261 mus musculu
8	133.5	3.1	5327	1 AC7_MOUSE	O9qx20 mus musculu
9	131	3.0	1805	1 HMW2_MYCGE	P47460 mycoplasma
10	131	3.0	2683	1 CENE_HUMAN	Q02224 homo sapien
11	130.5	3.0	2364	1 SPCO_HUMAN	Q01082 homo sapien
12	130	3.0	5430	1 AC7_HUMAN	Q9upn3 homo sapien
13	129.5	3.0	1312	1 RA50_YEAST	P12753 saccharomyc
14	129.5	3.0	1957	1 YD86_SCHPO	Q10411 schizosacch
15	128.5	2.9	1940	1 MYH3_HUMAN	P11055 homo sapien
16	128	2.9	3660	1 DMD_CHICK	P11533 gallus gall
17	128	2.9	3678	1 DMD_MOUSE	P11531 mus musculu
18	127.5	2.9	284	1 TPM1_SCHMA	P42637 schistosoma
19	127.5	2.9	1939	1 MYH1_HUMAN	P12882 homo sapien
20	127	2.9	1938	1 MYSS_CHICK	P13538 gallus gall
21	126	2.9	2230	1 GOGA_HUMAN	Q13439 homo sapien
22	126	2.9	4466	1 DYHC_ANTCR	P39057 anthocidari
23	125.5	2.9	1940	1 MYH3_RAT	P12847 rattus norv
24	125.5	2.9	2349	1 TPR_HUMAN	P12270 homo sapien
25	125.5	2.9	4466	1 DYHC_TRIGR	P23098 tripneustes
26	125	2.9	1829	1 DPOL_THEST	O33845 thermococcu
27	124.5	2.8	2470	1 TOR1_YEAST	P35169 saccharomyc
28	124	2.8	1679	1 YIO9_YEAST	P40457 saccharomyc
29	123	2.8	1044	1 YAF3_SCHPO	O09857 schizosacch
30	123	2.8	4563	1 APB_HUMAN	P04114 homo sapien
31	122.5	2.8	1453	1 Y373_BOVIN	O9tu23 bos taurus
32	122.5	2.8	1755	1 PEPL_MOUSE	Q9r269 mus musculu
33	122.5	2.8	1939	1 MYH4_HUMAN	Q9y623 homo sapien

34	122	2.8	2116	1 MYS2_DICDI	P08799 dictyostell
35	121.5	2.8	1333	1 CC25_CANAL	P43069 candida alb
36	121	2.8	283	1 TFMU_LOEMI	P31816 locusta mig
37	121	2.8	1938	1 MYHM_HUMAN	Q9ukx3 homo sapien
38	121	2.8	1959	1 MYH9_CHICK	P14105 gallus gall
39	121	2.8	2469	1 TEGU_HSVSA	Q01056 herpesvirs
40	120.5	2.8	2210	1 RPO_LYCAVA	P14240 lymphocytic
41	120	2.7	499	1 C72L_ARATH	O9ltm2 arabidopsis
42	119.5	2.7	1139	1 NGAP_HUMAN	Q9ujf2 homo sapien
43	119.5	2.7	1940	1 MYH3_CHICK	P02565 gallus gall
44	119	2.7	946	1 FTSK_CAJJE	Q46089 campylobact
45	119	2.7	1790	1 USOL_YEAST	P25386 saccharomyc

ALIGNMENTS

RESULT 1
CB80_HUMAN
ID CB80_HUMAN STANDARD; PRT; 790 AA.
AC Q09161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 80 kDa nuclear cap binding protein (NCBP 80 kDa subunit) (CBP80).
GN NCBP1 OR NCBP OR CBP80.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 119-128; 513-522 AND 687-693.
RX MEDLINE=94349369; PubMed=8069914;
RA Izaurralde E., Lewis J., McGuigan C., Jankowska E.,
RA Darzynkiewicz E., Mattaj J.W.;
RT "A nuclear cap binding protein complex involved in pre-mRNA
RT splicing.";
RL Cell 78:657-668(1994).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95023141; PubMed=7937105;
RA Kataoka N., Ohno M., Kangawa K., Tokoro Y., Shimura Y.;
RT "Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap
RT binding protein.";
RL Nucleic Acids Res. 22:3861-3865(1994).
CC -!- FUNCTION: INVOLVED IN MEDIATING U SNRNP EXPORT FROM THE NUCLEUS.
CC BINDS TO 5' CAPPED MRNA.
CC -!- SUBUNIT: THE CAP-BINDING PROTEIN (CBC) COMPLEX IS AN HETERODIMER
CC OF CBP80 AND CBP20.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: SOME TO YEAST GCR3/STO1 PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80030; CAA56334.1; -;
DR EMBL; D32002; BAA06769.1; -;
DR MM; 600469; -;
DR InterPro; IPR003890; EIF4G_cent.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
KW Nuclear protein; RNA-binding.
FT DOMAIN 3 20 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 790 AA; 91839 MW; F10DE7B9D16FDA0B CRC64;

Query Match 18.6%; Score 813; DB 1; Length 790;
Best Local Similarity 26.6%; Pred. No. 2.6e-42;

[illegible]

```

Db 196 YNTLLNIPYLFNFFNNRNDGLRTKVEE-----LLAVEQNYLVKTTDINLLREYNCE 247
QY 207 FESSLAEE-----KDFVEDL-----LDRIQ 225
Db 248 PPEYKVELVVRVLPVKKALLNNLQNLFPDMNHLTPOTDGEFNDALFLPSVDDLK 307
QY 226 S---LASNGWKLESV-PRPHLSFEAQL--VAGKFHELPRIPKMEQPPSDHSRAYSGKO 279
Db 308 SFVRLNKNFGSVDSMKWTPRFAHYVLPNSAGNFTVVPIS-----TYAGOL 354
QY 280 KHDALTRYQRRIRRNIPANKMEDVQPIDRFVWEYLLDVLFLNGCRKE-----C 331
Db 355 FNDII-----IDLVESEFNKE-----VAQVITLDFFRAGIFTEPGESIAQLI 400
QY 332 ASYWAN-LPVTFREYLMAEFLFSQILLP--QPPEFKLYTYTLVMDLCKALPGAPPAV 388
Db 401 ATYEENPLAFTFKIEDLAETILGLIFKLPVSQPF-AFYFTLLV-DICNSPKAIAPVF 458
QY 389 AGAVRALFEKISDLDMESRTRILWFSHLSNFOFIWPEWAFVLDLPKWA-----PKR 443
Db 459 GRAFRFFYSHDSLDLFELKLYLWFSIQMSNFSWKNWEMED--DSIKFGKYFNPVKV 516
QY 444 VFVEIILQREVLSYWDKIKOSIENATALELLPPK-----479
Db 517 NFAKNLIQKELRLT-----SNFSEVEDSLPOEFTKYLDTSVIPDQLINYYQSLEF 567
QY 480 -----AGNFMYSLEEG--KEKTEBQQL-----SABLSRKVKQEKOPARDIMVWIERTI 525
Db 568 GYTVEEDSVRKNDLYFROEGVPMENTVRKILDYTHKANSREVTLES-----ILGEL 620
QY 526 YPVHGFVET-----LTVQTLIDIGSKSETH-----LVTVLRYGVFSKLCPPD-NKQ 574
Db 621 KNEYGSIIDNRFVILLVQAVTDGSGRSLSHANKYINDLDEDLKTIKAKIELDIETKE 680
QY 575 VMLLSQVSTYKNNVQMTAVADRMMGYRLVSNQAIVRWVFSF-----617
Db 681 YIIIEAVLTFNANPQTGFLVADAFKYAGLTSRTIFFINETGLKNGNGLIETAEAV 740
QY 618 -ENVDDQHVSDQPEIIGNALNKYNYRISDRKISNTTKNVVAEKAASANARVEAAE 676
Db 741 FRNLQO-QISEE--NESCNNFEFVERLC-----TIANSTIDLL-----DVNADEIDIE--- 785
QY 677 SKLSLVEGE 685
Db 786 --IPKVNGE 792

RESULT 3
RA50_PVRHO
ID RA50_PVRHO STANDARD: PRT: 879 AA.
AC Q36887;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 ATPass.
GN RAD50 OR PH0929.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT *Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.*;
RL DNA Res. 5:55-76(1998).
CC -I- FUNCTION: Involved in DNA double-strand break repair (DSBR). The

```

```

CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -I- SUBUNIT: Forms a complex with mrell (By similarity).
CC -I- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP000004; BAA30025.1; .
CC InterPro: IPR003439; ABC transporter.
CC InterPro: IPR001687; ATE_GTP_A.
CC InterPro: IPR001238; RecF.
CC InterPro: IPR003405; SMC_C.
CC Pfam; PF00470; RecF; 1.
CC Pfam; PF02483; SMC_C; 1.
CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 30 37 ATP (BY SIMILARITY).
CC FT DOMAIN 141 744 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 879 AA; 103673 MW; FD4E30FD1BBCDB29 CRC64;

Query Match 3.8%; Score 166; DB 1; Length 879;
Best Local Similarity 19.3%; Pred. No. 0.013;
Matches 106; Conservative 105; Mismatches 191; Indels 146; Gaps 25;

QY 213 EKDFVEDLDRITQSLASNGWKLESVPR----PHLSFEAQLVAGKFHELPRIPKMEQSP 267
Db 202 EKSFTF-VLNEIRNISSN-----LPRLRLEGIEKEVKTLEATFNSITELAL----- 248
QY 268 PSDHSRAYSGKQKHDALTRYQRRIRRNIPANKMEDVQPIDRFVWEYLLDVLFLNGC 327
Db 249 ---RLGELNGKK-----GRLEIRIQLGSEIGIEKKRKSKELEEVKRE-----LPEL 291
QY 328 RKECASYMANLPVTRFYEYLMAEFLFSQILLPQPFPKTYLYTLVIMDLCKALPGAPPAV 387
Db 292 EKKETEY--RLIEFADFELVRKN-----ELEKRL-----GI 321
QY 388 VAGAVRALFEKISDLDMESRTRILWFSHLSNFOFIWPEWAFVLDLPKAPKRVFVQ 447
Db 322 LSNRLQEVKRIK--DAESKVARIRWIERLKEIQ-----E 355
QY 448 EILQREVRLSYWD---KIKQSTEN-ATALEELLPPKAGNFMYSLEEGKEKTEEQ----- 498
Db 356 KIMKLEPRVREPEDAMRLKAQMESLKSGLGLEPEKINEKLLY-LENRRKKELEEDKIT 414
QY 499 QLSAELSRKVKQKOTARDIMVWIEET--IYPVHGFVETLTIIVQTLIDIGSKSFTHLTV 556
Db 415 RKIGELNQRSDKRLA---IIEELKARGKCPVCGRELT-----EELKADL 456
QY 557 LERYGVQVFSKLPDNDKQVMLLSQVSTYKNNVQMTAVADRMMGYRLVSNQAI-VRWVF 615
Db 457 LRKYSLELSSI---EKEIQEAKALERQLRAEFKRVENELSRLLSLTKTIADQIIEIRL 512
QY 616 SPENVDDQHVSDQPEIIGNALNKYNYRISDRKISNIT--KNVLVAEKASANARVELE 673
Db 513 SKINLEDLRDKEEYELLKSESNKLKGEVLSKKEVKNELNDYKN-----ESTPKLEIJD 566
QY 674 RAESKLSLVE-----GEPLGLENPAKMKRL-----KSTVEKTGEAL-SLRESLEAKEA 721
Db 567 KAKKELSEIEDRLRLGLFTKIDELSGRIELEKFKHNYTEAKNAEKELDILESLEKDERE 626
QY 722 LINRALSE 729
Db 627 ELDKAFEE 634

```



```
RESULT 4
MY33_HYDAT STANDARD; PRT; 539 AA.
AC P39922;
DT 01-FEB-1995 (Rel. 31, Last Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYosin heavy chain, clone 203 (Fragment).
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakano M.Y., Stidwill R.P.;
RL submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC -!- ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST,
CC TO OTHER NON MUSCLE MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L35595; AAA29216.1; --
CC HSP: P24733; LWDC.
CC InterPro: IPR002928; Myosin_tail.
CC Pfam: PF01576; Myosin_tail; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding.
CC FT NON_TER 1 1
CC FT DOMAIN <1 38 GLOBULAR HEAD.
CC FT DOMAIN 39 >539 RODLIKE TAIL.
CC FT DOMAIN 39 520 COILED COIL (POTENTIAL).
CC FT NON_TER 539 539
CC SEQUENCE 539 AA; 62319 MW; 9CSAD5664060939D CRC64;

Query Match 3.2%; Score 140.5; DB 1; Length 539;
Best Local Similarity 23.2%; Pred. No. 0.23;
Matches 96; Conservative 60; Mismatches 156; Indels 101; Gaps 19;

QY 442 KRVFQEI-----LQREVR-----LSYWDKIKQSIENATALEELPPKAGPNMYSLEEGKE 493
DB 5 KKVAAQRIGLAVIQNRVRYKFLRHWSWKL-----YTKVQPLLVSARAEDEMRAKEELE 60
QY 494 KTEEQ-QLSALSRKVKEX-----QTARDMIVMIEETIYPVHGFEVTLTI 537
DB 61 AAQELKKDAEAKKMEELTEAMAQEKLYASLQAEITDRLTIEDKL-----L 109
QY 538 VVQTLIDIGSKSPT-----HLVTVL-ERYGOVESKCLPDNDKQVMLLSQVSTYWK 586
DB 110 NLQTVKDKLESSLNEALEKLDGEEHSLVLEEKIQAEAEKIDLTEKTEELQSNISRL-E 168
QY 587 NNQMTAVADRMMGYRLVSNQAIWR--VFSPENVDFHVSQ-----QFWEILGN 635
DB 169 TEKQNRDKQIDTL-----NEDIRKQDETISKMNAEKKHVDDELKDRTELQAAEDKCN 221
QY 636 ALNKTYNR-----ISDLRKDISNITKNVLAEKASANARVELEAEKSLVGEPEVLGN 691
DB 222 NLNKTNKLLESSIRETEQDLKKKEKSKMKLEKKEKVESDLKDNDRDKLSETE----- 273
QY 692 PAKMKRLKSTVEKTGAEKLSL-----RESLEAKALLNRLSETVELLLLLFQSLGLVL 745
DB 274 ----TLKETQDVLTVTRKESISDLNNAKGLQESQISQOLQKQIE-----LLAKTEE-----L 321
QY 746 KERLPDPTKVRSDQLKLSIGAEDDKPSAMD-VDSSENGNPKKCEVG-EREQMC 796
```

```
Db 322 EELENERKLRKSELORKELESRIEELQDOLETAGGATSAQVEVGKKREAPC 374
RESULT 5
RASO_ARCFU STANDARD; PRT; 886 AA.
ID RASO_ARCFU
AC O29230;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR AFI032.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (by
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001032; AAB90211.1; --
CC TIGR: AFI032;
CC InterPro: IPR003439; ABC_transportr.
CC InterPro: IPR001238; RecF.
CC InterPro: IPR002017; Spectrin.
CC Pfam: PF00470; RecF; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 31 38 ATP (BY SIMILARITY).
CC FT DOMAIN 148 728 COILED COIL (POTENTIAL).
CC SEQUENCE 886 AA; 103633 MW; D35641D499AAB58 CRC64;

Query Match 3.1%; Score 134.5; DB 1; Length 886;
Best Local Similarity 19.0%; Pred. No. 1.1;
Matches 106; Conservative 107; Mismatches 176; Indels 169; Gaps 25;

QY 376 LKALGCAFFPAVAGAVRALFEKISDL--DMESRTRLLWFHSHLSNFQITWPEEWAFFV 433
DB 114 LCFA-----HVTGTGIYVRQGEISDIRDESRIIRQIT-RIEDYEN--ANKNLGAV 164
QY 434 LDLPKWAPKRVFVQEIL-----QREVLSYWDKIKQSIENATALEELPPKAGPNFM 485
```

```

165 IRMLERKER--LKEFLSQEQIKQKEKAEIERISEEIKSIESLREKLESE-----V 217
QY 486 YSLEGEKTEKTEQOQLSALSRLVKVKEQARDMIVWIEETIPVGHFVTLTIIVOTLIDI 545
Db 218 RNLESRLKEHEKHSRLSLR--KOESSVLOEVRLGLEKL-----RELEKQKEVVERIEDL 272
QY 546 --GKSFTHLVTLERYGOVSKLCPDNDKQVM-----LLSQVSTYWKNN 588
Db 273 EKKAKEVLPKAEY--SILEKLLSEINQALRDVEKREGDLTRAAIGIQAOLKKAEDN 331
QY 589 VOMTAVA-----IDRMGYRLVSNQAIVRWVFSPEVDQ 622
Db 332 SKLEEITKRIELERELEFEKSHRLLETLKPKMDRMQIGAKLEKNL---TPDKVE- 386
QY 623 FHVSDQWEIIGNALNKTYNISDLRKDISNITKNV--LVAKASANAR-----VEL 672
Db 387 -----KNYDLSRAKEEKEETTEKLLKLIKAKSLKTRGAQLKKAVEEL 430
QY 673 EAAE-----SKLSLVEGEPVLGNPAKMKRLKSTVEKTGEAELSRESLEAKALLNR 725
Db 431 KSAERTCPVCGRELDDEEHRKKNIMAEYTRMKRIAEELAKADEIEKKLERLEKVB---K 486
QY 726 ALSTEVELLLLQFSLGVLKRLPDPKTVRSVDLKGIGAEDDKPSAMDVDSENGPKK 785
Db 487 ALEKQET-----VLKVR-----QWDELKAL--ENELSHDAEKLKSAESEE 525
QY 786 SCEVGER-----EQWCLSTGLVLTFT-----ROYASEIWP 817
Db 526 YRKVKERLDGLRGQOKILLSSASRIKELKSLREIEEALKNVESBRGELHRKIREGPE 585
QY 818 MEKLESEVFCEDVHPLF 835
Db 586 LEELEREV---OSLRPFY 600

RESULT 6
HMW2_MYCPN
ID HMW2_MYCPN STANDARD; PRT; 1818 AA.
AC P75471;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytoadherence high molecular weight protein 2 (Cytoadherence accessory
protein 2).
GN HMW2 OR MPN310 OR MP526.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97252497; PubMed=9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA Herrmann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT pneumoniae cytoskeletal protein HMW2 and cytoadherence.";
RL J. Bacteriol. 179:2668-2677(1997).
CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERESIN PROTEINS
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY

```

```

CC -----SIMILARITY)-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000051; AAB96174.1; -
CC EMBL; U59896; AAB52527.1; -
CC DR HSSP; P04083; 1B09.
CC Cytoadherence: Structural protein; Coiled coil; Complete proteome.
KW DOMAIN 31 880 COILED COIL (POTENTIAL).
FT DOMAIN 919 1607 COILED COIL (POTENTIAL).
FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
FT DOMAIN 1786 1817 COILED COIL (POTENTIAL).
SQ SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0FCBC0 CRC64;

Query Match 3.1%; Score 134.5; DB 1; Length 1818;
Best Local Similarity 19.1%; Pred. No. 3;
Matches 158; Conservative 104; Mismatches 256; Indels 309; Gaps 34;

QY 121 ASLVVVFETLLSSAATTVDDEKGNFSPQADFYVICILSSLPWGSSELAEOVPDEIERV 180
Db 235 AELQNYVETIAQNA-----NFOQCDAY-----WAQLKQVEOQIOTTKE 275
QY 181 LVGICAYLSIRKNSSTGLNFFHNG-----EFESLAEKDFV--EDLDRIOQLASNGW 232
Db 276 LVDEESTLKVRLNDA---DFYINSRLAELDLTSKINERDFVSKQAQDVKASLANLTK 331
QY 233 KLESYRPHLSFEAQLVAGKPHLRIKCMQEPSPSDHSRAYSGKOKHDALTRYPORIR 292
Db 332 EKERLSAEKDSPE-----
QY 293 RLNIFFPANKMEDVQPI-----DRFVVEEYLLDVLFLNGCRKCEKASYMANLPVTRYE 345
Db 348 NTALNDINRMEQENALFAKHLEQQOYEFERKQOESLLKLETHKQLOKQIG-----E 399
QY 346 YLMAETLFSQILLPQPPFKTYLYTLVIMDLCKALPGAFVAVGAVRALFEK---ISDL 402
Db 400 PKIESEAKSEALLIQE-----RELLEKREIDDL 428
QY 403 -----DMESRTRLILWFSH-----HLNSFOFIWPMWEAFVLDLPKWPKRKFVQEI 449
Db 429 LTQASLEVEQQRRTNQLKKEKHQVQHFON-----LVHAKKKLDQKRHYLAE- 476
QY 450 LQREVRLSYWDKIKOSIENATALELLPPKAGPNWYSLEEGKEKTEE-----QQ 499
Db 477 -OKRIDEQEIFKLKEKI--ATERREL-----EKLVLVKKQKQDOKDNDLLIFEKOLRQ 526
QY 500 LSAELSRVKVEKOTARDMIVWIEETIYPVGHFEVTLTIIVVOTLDTGKSFTHLVTVLER 559
Db 527 YQADFENIEEKQNE-----LPASOKSL---QKSFTOLKNKEAE 562
QY 560 YQGVFSKLCPD-----NDKQVML-----LSQVSTYWKNNVQMTAVA 595
Db 563 LNQAQKIAEDWAHLKQNHKHHADLEIFLEGEFNHLQEKHLLEARTQFDRVLSLSAR 622
QY 596 IDRMGYRLVSNQAIVRWVFSPEVDQFIVSQPW-----EILGNALNKTYNRI 644
Db 623 FKQKQA-ELVKQKQSLQTLTAFAFNKEQEAIV-ERDMKDRLANLEKQKEMLG-----DKV 673
QY 645 SDLRKDINSITKNVLVAEKASANARVELAAESKLSLVEGEPVLGNPAKMKRLKSTVEK 704
Db 674 HOFDENSINISKKLAERLAIKFEKELEPAQOKLSLD-----NNNNA---GLKLQLODK 724
QY 705 TGEAELSRESLEA-KEALLN-----
Db 725 LSESLKTERLELEAKSERILDPYDESSRRRIADYESDLQARLAEVKTLEKNQOETAAKSER 784

```



```
Db 546 EMKVLSSQYXKHL---LGVEDLLQKHALVEADIAIOAERVRGVNASQAQFATDGEYK 602
Qy 396 -----PEKISDLDMESRTR-----ILWFSHLSNFOFIWPEE---WA 431
Db 603 PCDPQVIRDRVAHMEFCYQELCOLAAERARLEESRRLW-----RFFWEMAEDEGW- 653
Qy 432 FVLDPKWPARKRVFOELQREVLRYSD-----KIKQSTENATALEELPPKAGPFM 485
Db 654 -----IREKILSSDDYGDKDLTSVMRLLSKSHRAFEDMSGRSG-HFE 695
Qy 486 YSLEEGKEKTEEQSLAELSRKREK-----QTARDMTVWTEETIYPVHGFEV 533
Db 696 QAIKEGEDMIAEEHFGSE---KIRERIYIREQWANLEQLSAIRKLEASL-LHQQA 751
Qy 534 TLTIVVQTLDI-----GSKSTH-----LVTVLRYGVQVFSK----- 567
Db 752 DADDIDAMMLDLIKLVSSNDVGHDEYTSQSLVKKHKDAVEITNYRPTIDTLHEQASALP 811
Qy 568 -----CPDNDKQVMLLSQV-----STYW 585
Db 812 QAHAESPDVKRGLAGIERCKEMAELTRKQALRDLTLALYKMFSEADACELWIDKEQW 871
Qy 586 KNNYQM-----TAVAIIDRMGVRLVSN----- 607
Db 872 LNNQIPEKLEDLVIOHREFSELEPMNNQASRVAVVQIARQLMHNCHPSEKEIRAQOD 931
Qy 608 QAIYRW-----VSPENVDFHVSQDPWEILGNALNKTYNRIISDLRK----- 649
Db 932 KLNTRWSQFRELVDKDKALLSALSQNYHLE-----CNETKSWIREKTRVIEST 981
Qy 650 -DISNITKNVLVAEKASANARVELEAESKLSLV---EGEPVLGENPKMKRLKLSVEXTG 706
Db 982 QDLGNDLAGVNALQRLKTMGERDLVAIEAKLSLDQKEAEKLESEHPDQQAILSLRAIS 1041
Qy 707 EAELSLRESLEAKALLNRALSETVLLLL---FOSFLGVLRKLPDPTKVRVQDLKSI 764
Db 1042 DWYEMKTTLNREASLGEA-SKLOQFLRLDLDLDFOSWLSRTQTAI----- 1085
Qy 765 GAEDDKPSAM-----DVSSENGNPKKSEVEGEQWCLSTLGLYLA---F 806
Db 1086 -ASEDMPNTLTAEEKLLTQHENIKNEIDYEDYQKMRDMGE-----MVTQGTQDAQYMF 1139
Qy 807 TRQYASEI---WPHMEKL---ESEVFGSDVHPLEPLQAISSALQF 845
Db 1140 LRQRLQALDTGWNELHKWENRQNLSSQSHAYQQFLRDTKQAEAF 1184

RESULT 8
ACF7_MOUSE
ID ACF7_MOUSE STANDARD; PRT; 5327 AA.
AC Q90X20; P97394; P97395; P97396;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (Microtubule actin crosslinking
factor) (MACF).
GN ACF7 OR ACP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN-BALB/C;
RX MEDLINE-20069791; PubMed-10601340;
RA Leung C.L., Sun D., Zheng M., Knowles D.R., Liem R.K.H.;
RT "Microtubule actin cross-linking factor (MACF): a hybrid of dystonin
and dystrophin that can interact with the actin and microtubule
cytoskeletons."
RL J. Cell Biol. 147:1275-1286(1999).
RN [2]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC STRAIN-BALB/C; TISSUE-Brain;
```

MEDLINE-97124842; PubMed-8954775;
Bernier G., Mathieu M., De Repentigny Y., Vidal S.M., Kothary R.;
"Cloning and characterization of mouse ACF7, a novel member of the
dystonin subfamily of actin binding proteins.";
Genomics 38:19-29(1996).
-1- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
MICROTUBULES.
-1- SUBCELLULAR LOCATION: Cytoplasmic.
-1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1, 2 (SHOWN HERE) AND 3; ARE
PRODUCED BY ALTERNATIVE SPLICING.
-1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN LUNG, BRAIN, SPINAL CORD,
SKELETAL AND CARDIAC MUSCLE, AND SKIN.
-1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
-1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
-1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
-1- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF150755; AAD32244.1; -;
EMBL; U67203; AAC52988.1; -;
EMBL; U67204; AAC52989.1; -;
EMBL; U67205; AAC52990.1; -;
HSPF; Q01082; 1BRR.
MGI; MGI:108359; Aclp7.
InterPro; IPR001589; Actinin_act_bind.
InterPro; IPR001715; Calponin_hom.
InterPro; IPR002048; EF-hand.
InterPro; IPR003108; GAS2.
InterPro; IPR001452; SH3.
InterPro; IPR002017; Spectrin.
Pfam; PF00307; CH; 2.
Pfam; PF00036; ehand; 2.
Pfam; PF02187; GAS2; 1.
Pfam; PF00018; SH3; 1.
Pfam; PF00435; spectrin; 37.
SMART; SM00033; CH; 2.
SMART; SM00054; EFh; 2.
SMART; SM00243; GAS2; 1.
SMART; SM00326; SH3; 1.
SMART; SM00150; SPEC; 32.
PROSITE; PS00019; ACTININ_1; 1.
PROSITE; PS00020; ACTININ_2; FALSE_NEG.
PROSITE; PS00021; CH; 2.
PROSITE; PS00002; SH3; FALSE_NEG.
PROSITE; PS00018; EF_HAND; 2.
Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain;
Alternative splicing.
DOMAIN 1 295
DOMAIN 78 181
DOMAIN 194 295
DOMAIN 314 355
REPEAT 591 623
REPEAT 680 784
REPEAT 786 800
REPEAT 871 923
REPEAT 1250 1272
REPEAT 1287 1342
REPEAT 1458 1534
REPEAT 1593 1660
REPEAT 1817 1886
REPEAT 1934 2044
REPEAT 2262 2282
REPEAT 2376 2397
REPEAT 2400 2509
ACTIN-BINDING (BY SIMILARITY).
CH 1.
CH 2.
SPECTRIN 1.
SPECTRIN 2.
SPECTRIN 3.
SPECTRIN 4.
SH3.
SPECTRIN 5.
SPECTRIN 6.
SPECTRIN 7.
SPECTRIN 8.
SPECTRIN 9.
SPECTRIN 10.
SPECTRIN 11.
SPECTRIN 12.
SPECTRIN 13.

RT "The minimal gene complement of Mycoplasma genitalium.";
 RL Science 270:397-403(1995).
 RN [2]
 RC SEQUENCE OF 557-659 FROM N.A.
 RP STRAIN-ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing.";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -|- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS.
 CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U39701; AAC71437.1; -;
 DR EMBL; U02165; AD12447.1; -;
 DR TIGR; MG218; -;
 DR InterPro; IPR003364; Seryl_trna_N.
 DR Pfam; PF02403; Seryl_trna_N; 1.
 KW Cytadherence; Structural protein; Coiled coil; Complete proteome.
 FT DOMAIN 28 838 COILED COIL (POTENTIAL).
 FT DOMAIN 914 1591 COILED COIL (POTENTIAL).
 FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).
 FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;

Query Match 3.08; Score 131; DB 1; Length 1805;
 Best Local Similarity 20.18; Pred. No. 4.9;
 Matches 176; Conservative 139; Mismatches 302; Indels 258; Gaps 41;

QY 49 FLQCAEQHPKIPLYGLIGLLENDEFQKLVESYHANFOVALDSGNSIRILLRF 108
 DB 192 YLLNVLDQYLNELQLENQKLLSLEYENTYRELVSADNELQNVYENIDQIQFKQY 250
 QY 109 MT-----SLCSKVIQPSLIWFFETLLSSAATTVEEKGNPSWQPOADFVVICILSSLP 163
 DB 251 QTYRDELSOLERKIQLTKQELVDKE---SALRVKIDD-----ADFYINARLAELD 297
 QY 164 WGGSELA-----EQVDEIERVLVGIOAYLSIRKNSSTSGLNFHNGEFPSSLAEKDFV 217
 DB 298 DVAKQLSPQDQITKQNAQHVDEKLVAL-----NKEKDRLNTQKEAFF--NLRSQALI 347
 QY 218 EDLLDRIQSLASNGWKLESVPRPHLSFPAQLVAGKFHELPIKMCQEPSPPSDHSGRAYSG 277
 DB 348 D--INKLQO-----ENELFAKHL-----EQQNEFFE 371
 QY 278 KOKHDAL----TRYPQIRRLNIF---PANKMEDYQPIDRFVVEYLLDVLFLNGCRKE 330
 DB 372 QKQSDSLKLETKYKALQHKINEKFNESATKSEELLNQRLELFK-----RRE 419
 QY 331 CASYMANLPVTRFEYLAETLFSQILLPQ-----PPKTLTYTLVINDLCKALPGAPPA 386
 DB 420 IDTLITO--ASLEYEH---QRESSOLLKDKQNEVQHFQNLQYAKKELDKERNLLDQOKK 474
 QY 387 VVAGAVRALFKISDLDMESTRLTLWFSHLSNPFQIWPWEAWFVLDLPKWPAPKRVFV 446
 DB 475 VDSEAFOLKKEVA-----QERKL-----EELYLV-----KK--- 502
 QY 447 QEILQREVRLSYWDK-IQO-----SIEN---ATALEELPPKAGNFMYSLSEEGKEK---TEE 497
 DB 503 QKQDQKENELFFFEKQLQKHQADFENELEAKQOQLFEAKHALERSFIKLEKDEKDLNTKA 562

QY 498 QQLSAELSRKVKQKOTARDMIVMIETIYPVHGFEVTLTVVQTLDDIGSKSFHLVTVL 557
 DB 563 QQIANEFSQLTKDKSAD-----FELMLQNEVENLQOEKQKLFQER-TYF 607
 QY 558 ERYGVFSKLPDNDKQVMLLSQVSTYKNNVOMTAVADRM-----GYRLVSNQAIVRW 613
 DB 608 ERNAAVLSNRL--QOKREBLOQKET-----LDQLTKSFEQERLINQREHKEL 653
 QY 614 VPSPENVQFHVSDQPWELTGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELE 673
 DB 654 VASVE-----KOKETLG-----KKLQDFQSTSLNASKLAERMAIKFKEKE 697
 QY 674 AAESKL-----SLVEGEPVLGNPAKMRKL-----STVEKTEGEAELSRL 713
 DB 698 ATEKOLLNDVNAEVIQADLAQLNOSLQERSELQNAKQRIADFHNDLSKLKLEYELSLQ 757
 QY 714 ESLEAKEAL-----LNRLALSETV-LLLLLFQSLGV-----LKERLPDPT 753
 DB 758 KRLQELQTLLEANOKQHSYQNAVFEGLDKLNREKQAFNLRRKKQTMEDVDAIKORLSDKH 817
 QY 754 KVRVQV-----DLKS-----IGAEDDKFSAMD-----VDSNGNPKKSCVEGE 791
 DB 818 QALNMQQAELDRKTHLNNFLNHDADQKSLQDLATVKTOKLIDLE-----RSALLEK 872
 QY 792 REQWCLSLTGLYLTAFTRQVA--SEIWPHEKLESE 824
 DB 873 QREFAENVAGFKRHSWNKTSQLOKIYELTKKOESE 907

RESULT 10
 CENE_HUMAN
 ID CENE_HUMAN STANDARD; PRT: 2663 AA.
 AC Q02224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-2003 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENPE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and HUBB1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -|- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -|- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -|- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -|- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15005; CAA78727.1; --
DR PIR: S28261; S28261.
DR HSP: FI7119; 3KAR.
DR MIM: I17143; --
DR InterPro: IPR001752; kinesin.
DR Pfam: PF00225; kinesin; 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS00667; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
FT Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 3.0%; Score 131; DB 1; Length 2663;
Best Local Similarity 17.3%; Pred. No. 8.5;
Matches 195; Conservative 172; Mismatches 373; Indels 384; Gaps 43;

QY 17 EYGTSDYKDHETCTCGVI---RREIERSGDQVLPELLOCAEQLPKIPYGLIGLLNL 73
DB 327 KYAKNTPYNEVSTDEALKRYRKEIMDLKKQLEEVSLFTRAQMEKQDLAQL-----L 380

QY 74 ENEDFVOK-----LVESVHANFOVALDS-----GNCNSIR----- 103
DB 381 EEKDLQKQVONEKIENLTRLMTVSSLTQQELKAKRRKRVTWCLGKINKMKNVADQF 440

QY 104 -----ILLRPTSLICKSVQIPASLIWVETLSSAATVDEEKGPSWQ 148
DB 441 NIPTNTTKTKLSINLLREIDESVCE-----SDVFNLTDLTSEIENWPAWK 489

QY 149 PQADFVVICITSLSPGGSELA---EQVDEIERVLVGIQAYLSI-----RKNSSSTG 198
DB 490 LNQENIESELNLRADYNLDYDQLTEKEEMELKLEKNDLDEFALERKTKDQE 549

QY 199 LNFH-----NGEFESLAEKDFV-----EDLLDRIQSLASN----- 230
DB 550 MOLHIEISNLKVLKRVYVNDLENELSKVELLEKEDEQIKKLOEYIDSKLENIKMD 609

QY 231 -GWKLESVRP-----HLSFAQLVA-----GKFEHLRPIKMEQSPSPDHSRAVSGKQ 279
DB 610 LSYLSIESIEDPKQKQMTLEDATVETVALDRAKRESAFLRSENLEKMK-----ELATTYKQ 664

QY 280 KIDALTRYQIRRLNFPANKMEDVQ-----IDRFVVEYLLDVLFLNGCR 328
DB 665 MENDIQLOQLEAKKMKQVDELQSAFNEITKTLTSDIGKVPKDLL-----CN 715

QY 329 KECASYMANLPVTFRYEYLMAETLFSQILLPQPPKTYTLVIMDLCKALPG----- 382
DB 716 LELEGKITDLOKELNEVEENALREVILLSE-----LAKSLFSEVERLR 760

QY 383 -----APPAVW---AGAVRALFEKIS----- 400
DB 761 KEIQDKSEELHIITSEKDLKLFSEVHKESRVQGLLEEIGTKDKDLATQSNYKSTQDEQF 820

QY 401 -----DLDMESRRLILWTFSHHLSNQTQFWPWEENAFVLDLPKAPK-----R 443
DB 821 NFKTLHMDFEQYKVMLEENRMNQ-----EIVNLSKEAQKFDSSLGALKATELS 869

QY 444 VFQEIILQ---REV---RLSYWDKIKQSTEN-----ATALEELLPPKAGPNFMYSL 488

DB 870 YKTOELQKTRVOERLNEQKLEQLENRDPSLQTVREKTLITLTKL-----QOTL 921
QY 489 EEGKEKTEEQ-----QLSAELSRKV-----KEKOTARDMIVWIEETIY 526
DB 922 EEVKTTLTQEKDDKLQKQESLQIERDQLKSDIHTVNNNIDTQOLRWALESQKHQETIN 981
QY 527 P-----VHGFEVT-----LTIVVQTLIDIGSKSFTHLV-----VLER 559
DB 982 TLKSKISBEEVSRNLHMEENTGETKDEFOQKWGIDKKQDLEAKNTQTLTADVKDNEIEQ 1041
QY 560 YGVFSLKPCPDNDKQVLLS-----QVSTYWNKNVOMTAVAD--RMMGYRLVSNQIV 611
DB 1042 QRKIFSLIQEKNELQOMLESVIAEKQKLTDLKENIEMTIENQEEELRLGDELAKQOEIV 1101
QY 612 RWVFPSPNDVDFHSDPWEILGNALNKNTYRISDLRDKDISNITKNVLVAEKASANARVE 671
DB 1102 A-----QEKNHAIKKEG-----LSRTCRLAEVEEKKLEKSKSQQLQEQKQQLNQQEE 1149
QY 672 LEAAESKLSLVEGEPVLGENPARKRLKSTVEKTGEAELS----- 711
DB 1150 MSEWQKKINEIENL---KNELANKELTLEHMETERLELAQKLNENYEEVKSTIKERKVL 1205
QY 712 --LRESLEAKELNRLALSETV-----LLLLFQSLGVLKRLPDPT-KV 755
DB 1206 KELQKSFETERDHLRGYIRIEATGLQTKBELKIAHILKHEQHTIDELRRSVSEKTAQI 1265

QY 756 RSVDQLKS--IGAEDDKPSAMDVDSNGPNPKSCVCEGREQWCLSTLGYLTAFTQVASE 813
DB 1266 INPDDEKSTKTKQEEIPVLUHEEQELLPNVKKVSETQ-----TWNELELLTEQSTTK 1318
QY 814 IWPHEKLESEVF-----SGEDVHPLF-----LQAISSALQ 844
DB 1319 DSTTLARIEMERLRLNEKFOESQEEIKSLTKERDNLTKIKEALE 1362

RESULT 11
SPCO_HUMAN STANDARD; PRT; 2364 AA.
ID Q01082; Q16057; O60837;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)
DE (Beta-II spectrin) (Fodrin beta chain).
GN SPTBN1 OR SPTB2
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=92406787; PubMed=1527002;
RA Hu R.J., Watanabe M., Bennett V.;
RT "Characterization of human brain cDNA encoding the general isoform of
RT beta-spectrin".
RL J. Biol. Chem. 267:18715-18722(1992).
RN [2]
RP SEQUENCE OF 293-1544 FROM N.A.
RX MEDLINE=94010920; PubMed=8406479;
RA Chang J.G., Scarpa A., Eddy R.L., Byers M.G., Harris A.S.,
RA Morrow J.S., Watkins P., Shows T.B., Forget B.G.;
RT "Cloning of a portion of the chromosomal gene and cDNA for human beta-
RT fodrin, the nonerythroid form of beta-spectrin".
RL Genomics 17:287-293(1993).
RN [3]
RP SEQUENCE OF 2087-2168 FROM N.A. (SHORT ISOFORM).
RC TISSUE=Skeletal muscle;
RX MEDLINE=20267884; PubMed=10806113;
RA Hayes N.V.L., Scott C., Heerkens E., Ohanian V., Maggs A.M.,
RA Pinder J.C., Kordeli E., Baines A.J.;
RT "Identification of a novel C-terminal variant of beta1 spectrin: two

isoforms of betaII spectrin have distinct intracellular locations and activities.";

[4] J. Cell Sci. 113:2023-2034(2000).

RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 173-280.

RX MEDLINE-97307247; PubMed-9164454;

RA Carugo K.D., Banuelos S., Saraste M.;

RT "Crystal structure of a calponin homology domain.";

RL Nat. Struct. Biol. 4:175-179(1997).

RN [5]

RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 173-281.

RX MEDLINE-99036861; PubMed-9817844;

RA Banuelos S., Saraste M., Carugo K.D.;

RT "Structural comparisons of calponin homology domains: implications

for actin binding.";

RL Structure 6:1419-1431(1998).

CC -!- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,

CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS

CC THIS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE

CC CYTOSKELETON AT THE MEMBRANE.

CC -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS ARE

CC CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO TETRAMERS.

CC The short form can not bind to the axonal protein fodaxin.

CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a

CC short form; are produced by alternative splicing.

CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.

CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.

CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC -!- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

CC -----

DR EMBL; M96803; AAA60580.1; -

DR EMBL; S65762; AAB28324.1; -

DR EMBL; AJ005694; CAA06678.1; -

DR EMBL; AJ238723; CAB91088.1; -

DR PDB; 1AA2; 04-FEB-98.

DR PDB; 1BKR; 26-FEB-99.

DR MIM; 182790; -

DR InterPro; IPR001589; Actinin_act_bind.

DR InterPro; IPR001715; Calponin_hom.

DR InterPro; IPR001849; PH.

DR InterPro; IPR002017; Spectrin.

DR InterPro; IPR001605; Spectrin_ph.

DR Pfam; PF00307; CH; 2.

DR Pfam; PF00169; PH; 1.

DR Pfam; PF00435; spectrin; 17.

DR PRINTS; PR00683; SPECTRINPH.

DR SMART; SM00033; CH; 2.

DR SMART; SM00233; PH; 1.

DR SMART; SM00150; SPEC; 16.

DR PROSITE; PS00019; ACTININ_1; 1.

DR PROSITE; PS00020; ACTININ_2; 1.

DR PROSITE; PS50021; CH; 2.

DR PROSITE; PS50003; PH_DOMAIN; 1.

KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;

KW Alternative splicing; 3D-structure.

FT DOMAIN 1 275 ACTIN-BINDING (BY SIMILARITY).

FT DOMAIN 54 158 CH 1.

FT DOMAIN 173 275 CH 2.

FT REPEAT 276 384 SPECTRIN 1.

FT REPEAT 385 498 SPECTRIN 2.

FT REPEAT 499 608 SPECTRIN 3.

FT REPEAT 609 714 SPECTRIN 4.

FT REPEAT 715 819 SPECTRIN 5.

FT REPEAT 820 925 SPECTRIN 6.

FT REPEAT 926 1032 SPECTRIN 7.

FT REPEAT 1033 1139 SPECTRIN 8.

FT REPEAT 1140 1245 SPECTRIN 9.

FT REPEAT 1246 1350 SPECTRIN 10.

FT REPEAT 1351 1462 SPECTRIN 11.

FT REPEAT 1463 1562 SPECTRIN 12.

FT REPEAT 1563 1668 SPECTRIN 13.

FT REPEAT 1669 1775 SPECTRIN 14.

FT REPEAT 1776 1881 SPECTRIN 15.

FT REPEAT 1882 1987 SPECTRIN 16.

FT REPEAT 1988 2133 SPECTRIN 17.

FT DOMAIN 2197 2307 PH.

FT VARSPLIC 2141 2168

FT VARSPLIC 2169 2364

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

FT CONFLICT 1411 1411

Query Match 3.0%; Score 130.5; DB 1; Length 2364;
Best Local Similarity 17.4%; Pred. No 7.7; Mismatches 156; Indels 383; Gaps 47;
Matches 174; Conservative 156; Mismatches 156; Indels 383; Gaps 47;
SQ SEQUENCE 2364 AA; 274629 MW; 0122DB3DF25872CC CRC64;

QY 118 IOPASLIVVFETLLSSAATTVDEEKNPQWQADFFVICILSLPQGGSELAEQVPDEI 177
Db 298 IETEKIEKYESLASDLLEWIEQ-----TIILNNRKFANS----- 333
QY 178 ERLVGIQAYLSIRKNSSTSGLNFFNGEFESSLAEKDFVEDLLDRQS-LASNGWKLES 236
Db 334 ---LVGVQOOL-----QAFTYRTVEKPFTEKGNLEVLFTTQSKMRANQKV-Y 381
QY 237 VPRPHLSFQAQLVA-----GKFEHLRPIKCMQOPSPSDHSRAV--- 275
Db 382 MPR-----EGKLISDINKAWERLEKAEHERELALRNELIRQEKLEQARREDRAAMRET 436
QY 276 -----SGKQKHA-----LTRYQIRIRLNIFPANKME-----D 304
Db 437 WLSENQRLYSQDNFGFDLPAAVEAATKKHEAETDIAAYEERVOAV-VAVARELEAENVHD 495
QY 305 VOPI-----DRFV-VEEYLDVLVYNGCRKECASYANLPTVRYEYLMETLF----- 353
Db 496 IKRITARKONVIRLWEYLLLEL-----RARQ-----RLENNLGLQKIFQEMLYINDMMD 545
QY 354 -SQILLPQPPKTLV-----YTLVIMDL-----CKALPGAF----- 384
Db 546 EMKVLVLSQDYGKHLGVEDLLQKHTLVADIGIQIAERVRGVNASAKFATDGEYKPCD 605
QY 385 PAVVAGAVRAL---FEKISDLDMESRRL-----ILMFSHHLSNFQFIWPEE---WAFVL 434
Db 606 PQVIRDRVAHMEFCYQELCOLAAERARLESRRLM-----KPFEMAEPEGW----- 653
QY 435 DLPKWPAPRVFOEILQREVRLSYD-----KIKOSIENATALEELLPPKAGPNFYSL 488
Db 654 -----IREKEKILSSDDYDGKDLTSVMRLLSKHPAFEDEMSGRSG-HFEQAI 698
QY 489 EBGK-----EKTEE-----QQLSAELSRKVE-----KOTARDMIV 519
Db 699 KEGEDMAIEHFGSEKIRIRIIVQWANLEQLSAIRKRRLEASLLHOFQADADDIDA 758
QY 520 WIEETIYPVHGFEV-----TLTVVQTLTDIGSKSTHLVT----- 555
Db 759 WMLDILKIYSSSDVGHDYESTQSLVKKKHVDAAEIANRYPTLDTLHEQASALPOEAESP 818
QY 556 -----VLERYQGV-----FSKLCPDNDKQVLLSQVSTYKNNVQM 591
Db 819 DVGRLSGIEERYKEVAELTRLRKQLQDFTLALYKMFSEADACELWIDE-KEQWLNNOI 877
QY 592 -----TAVADRMMGYRLV-----SNQAIVRW 613
Db 878 PEKLEDLEVIQHRFESLEPEPMNNQASRVAVVNIQIARQLMHSGHPSEKIEKAQODKLNTRW 937
QY 614 -----VFSPENVDQFHVSDQPWEILGNALNKTNYRISDLRK-----DISNI 654
Db 938 SQFRELVDKDKDALLSALSIGNYHLE-----CNETKSWIRETKYESTQDGLND 987


```

QY 655 TKNVLAEKASANARVELEAESKLSLV--EGEPVLGENPAKMKRLKSTVEKTGEAELSL 712
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 988 LAGVMAIORLTKWGRDLVAIEAKSLDQEAKELESEHPDQAAILSLRAELSDVWVEEM 1047
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 713 RESLEAEKALLNRALETSEVILL--FQSFGLGVKLERLPDPTKVRSVQDLKSTGAEDDK 770
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1048 KTKLNREASLGEA--SKLOFLRDLDDFQSWLSRTQTAI-----ASEDM 1090
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 771 PSAM-----DVGSENGNPKKSCVEGREONCLSLPLGLYLA---FTROYAS 812
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 PNLTAEAKLLTOHENIKNEIDYEDYQKMRDGE-----WTQGTDAQYMFRLRQLQ 1145
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 813 EI---NPHMEKL---ESEVFSGEDVHPLFLQAISSALQF 845
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1146 ALDTGWNELHKWENRQNLSSQSHAYQQFLRDTKQAEAF 1184
|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
ACF7_HUMAN
ID ACF7_HUMAN STANDARD; PRT: 5430 AA.
AC Q9UPN3; Q9UKP0; Q9ULG9; Q9H540; O75053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Actin cross-linking family protein 7 (Macrophin) (Trabeculin-alpha)
DE (620 kDa actin-binding protein) (ABP620).
GN ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20001959; PubMed=10529403;
RA Okuda T., Matsuda S., Nakatsugawa S., Ichigotani Y., Iwahashi N.,
RA Takahashi M., Ishigaki T., Hamaguchi M.;
RT "Molecular cloning of macrophin, a human homologue of Drosophila
RT kakapo with a close structural similarity to plectin and dystrophin.";
RL Biochem. Biophys. Res. Commun. 264:568-574(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20026884; PubMed=10559237;
RA Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
RA Sutherland R., Salgia R., Griffin J.D., Perlman L.H., Chen L.B.;
RT "Molecular cloning and characterization of human trabeculin-alpha, a
RT giant protein defining a new family of actin-binding proteins.";
RL J. Biol. Chem. 274:33522-33530(1999).
RN [3]
RP SEQUENCE OF 868-2350 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirotsawa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [4]
RP SEQUENCE OF 1544-5057 FROM N.A.
RA Corby N.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 3734-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakaajima D., Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
CC -1- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO

```

```

CC CC MICROTUBULES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB029290; BAA83821.1; .
CC EMBL; AF141968; AAF06360.1; .
CC EMBL; AB033077; BAA86565.1; .
CC EMBL; AL137853; CAC15920.1; .
CC EMBL; AB007934; BAA32310.1; .
CC HSSP; Q01082; 1BKR.
CC InterPro: IPR001589; Actinin_act_bind.
CC InterPro: IPR001715; Calponin_hom.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR003108; GAS2.
CC InterPro: IPR001452; SH3.
CC InterPro: IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00036; efhand; 2.
CC Pfam; PF02187; GAS2; 1.
CC Pfam; PF00435; spectrin; 35.
CC SMART; SM00033; CH; 2.
CC SMART; SM00054; EFH; 2.
CC SMART; SM00243; GAS2; 1.
CC SMART; SM00150; SPEC; 35.
CC PROSITE; PS00019; ACTININ_1; 1.
CC PROSITE; PS00020; ACTININ_2; FALSE_NEG.
CC PROSITE; PS50021; CH; 2.
CC PROSITE; PS00018; EF_HAND; 2.
CC PROSITE; PS50002; SH3; FALSE_NEG.
CC Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.
CC DOMAIN 1 295
CC DOMAIN 78 181
CC DOMAIN 194 295
CC REPEAT 314 355
CC REPEAT 591 623
CC REPEAT 680 784
CC REPEAT 786 800
CC DOMAIN 871 923
CC REPEAT 1250 1272
CC REPEAT 1287 1342
CC REPEAT 1455 1534
CC REPEAT 1547 1659
CC REPEAT 1815 1891
CC REPEAT 1932 2042
CC REPEAT 2260 2280
CC REPEAT 2372 2395
CC REPEAT 2398 2507
CC REPEAT 2510 2618
CC REPEAT 2621 2728
CC REPEAT 2731 2838
CC REPEAT 2841 2945
CC REPEAT 2987 3024
CC REPEAT 3136 3163
CC REPEAT 3187 3274
CC REPEAT 3277 3383
CC REPEAT 3386 3492
CC REPEAT 3495 3601
CC REPEAT 3604 3673
CC REPEAT 3713 3819
CC REPEAT 3832 3927

```

```
FT REPEAT 3982 4043 SPECTRIN 27.
FT REPEAT 4046 4152 SPECTRIN 28.
FT REPEAT 4155 4262 SPECTRIN 29.
FT REPEAT 4265 4371 SPECTRIN 30.
FT REPEAT 4374 4481 SPECTRIN 31.
FT REPEAT 4484 4590 SPECTRIN 32.
FT REPEAT 4593 4700 SPECTRIN 33.
FT REPEAT 4707 4808 SPECTRIN 34.
FT REPEAT 4811 4917 SPECTRIN 35.
FT REPEAT 4920 5054 SPECTRIN 36.
FT REPEAT 5030 5054 SPECTRIN 37.
FT CA_BIND 5096 5107 EF-HAND 1 (POTENTIAL).
FT CA_BIND 5132 5143 EF-HAND 2 (POTENTIAL).
FT DOMAIN 5276 5283 POLY-SER.
FT DOMAIN 5355 5370 4 X 4 AA TANDEM REPEATS OF [GSI-S-R-[AR].
FT CONFLICT 1 MSSDEETLSRSCRSRSEYRSRSGSLSPCPG
FT FT TLPWNLPHQKRRKSDVLDPAERAVRV -> MFFVLW
FT FT AGIPGRDVSGLQPLPPGFKFCTASRVAVI (IN REF.
FT FT 2).
FT CONFLICT 575 594 VAISSEDEGNLRFVVELLS -> GPSAPLKMKAISDLCLM
FT FT YCL (IN REF. 1).
FT CONFLICT 1487 1487 A -> T (IN REF. 2 AND 3).
FT CONFLICT 1963 1963 V -> A (IN REF. 1).
FT CONFLICT 2052 2052 E -> D (IN REF. 1).
FT CONFLICT 2083 2083 E -> K (IN REF. 2).
FT CONFLICT 2290 2290 M -> V (IN REF. 2).
FT CONFLICT 2321 2321 C -> Y (IN REF. 1).
FT CONFLICT 2343 2363 MISSING (IN REF. 2).
FT CONFLICT 2344 2350 SILPSVG -> EYRLFKI (IN REF. 3).
FT CONFLICT 2523 2523 Q -> R (IN REF. 2 AND 4).
FT CONFLICT 4670 4670 S -> T (IN REF. 2).
FT CONFLICT 4833 4833 MISSING (IN REF. 2).
SQ SEQUENCE 5430 AA; 620346 MW; 91ADB7F7580B440B CRC64;

Query Match 3.0%; Score 130; DB 1; Length 5430;
Best Local Similarity 19.3%; Pred No. 27;
Matches 162; Conservative 134; Mismatches 339; Indels 204; Gaps 36;

QY 52 QCAEQL-PHKIP---LYGFLIGLNLLENED-----FVQKLVSFVHNFQALDSDGNCNSI 102
DB 1979 QLQEEALAEHVPVEXLQKQVARDIMETEGEPADHRHVEQTTDSLSHFQ-----SL 2029
QY 103 RILLRFTSLGSKVQIOPASLIIVFETLLSSAATTVEEKGPNQSQPADFYVICILSSL 162
DB 2030 SVSLAERSLLQKAQSQSVQESLESLLQSIGEVQNLGKQ-----VSSL 2076
QY 163 PMGSGELAEQVDEIERVVLVGIQAVLSIRKNSSTGLNFFHNGEFESSLAEDFVBDLLD 222
DB 2077 SSGVIOEALATNKLQDIAROKSSLEATREMT---RPMETADSTTAAVLOGLAEVSQ 2133
QY 223 RIQSLASNGWKLES-----VPRPHLSFEAQLVAGKPHLRPIKCMQPSPPSDHSRAYS 277
DB 2134 RFEQLCLOOQKESLSIKLLPOAEM-FEH--LSGKLOQF-----MENKS-----RMLASG 2180
QY 278 KQKHDAIYRPIRRLNIFPANKMEDVQIDRFVE-----EYLLDVLEY---LNCCKRE 330
DB 2181 NOPQODITHFOQIOELNLEMDQENLDTLEHLVTELSSCGFALDLCQHQDRVQNLKRD 2240
QY 331 CASYMANLPVTRYEYVMAETLFSQILLPQPFPFKLYVTLVIMDLCKALPGAFPAVAVAG 390
DB 2241 ----FTELQKTVKEREKDASSCOEQL-----DEFKRLVT--FOKWLKTEGSIPTET- 2288
QY 391 AVRALFEKISDLDMSRTRILWFSHLLNFQFIPWEEWAFVLDLPKWPARKVFVQEIL 450
DB 2289 -----SMSAKELEKQI-----EHLKSL-----LDDNASKGTLVEEIN 2320
QY 451 QREVL-----SYWDKIKQSIENATA-----LBEILPPKAGPNFMTS 487
DB 2321 CRGTSLENLIMEITAPDSQKTSILPSVSGSVGVNGYHTCKDLTEIQCDMSDVNLKYE 2380
QY 488 LEEGKEKTEEQALSALSRKVKRQKTARDMIWIEETIYFVHGFETLTIIVQTLDDIGS 547
```

```
DB 2381 KLGVLHERQESLQAILNRMEEVHKEANSVLQWLE-----SKEEVLKSDMAMSS 2429
QY 548 KSFTHLVTVLYRGQVFSKLPDNDQVMLLSQVSTYWKNNQVMTAVAIIDMMGYRLVSN 607
DB 2430 PTKTETVTKRAESNKAFLAELEQNSPKIQKVE-----ALA----- 2465
QY 608 QATVRWFSPENVDPQHVSDQPEWELIGNALNTKYNRISDLRKDISNTTKNVLVAEKASAN 667
DB 2466 GLLVTPNSQE-----AENWKIQEELNSRWEATEV-----TVARORQLEESASHL 2512
QY 668 ARVELEAAESKLS--LVEGEPVLGE-----NPAKMKRLKSTVE---KTGEAELSLRESL 716
DB 2513 AC--FQAAESQLQPLWMEKELMGMVLGPLSIDNMLNAQKQVQFMKLEFEARRQOHEQL 2570
QY 717 -BAKELL-----NRALSETEVLLLLLQSFGLGVKKE---LPDPTKVRVQDLKLSIGAE 768
DB 2571 NEAAQGLILPGDVSLSSTQV-----QKELQSIQKQWELTKLNSRSQIDQAIYKST 2624
QY 769 DKPSAMDVDSENGNPKKSCVEGER--EOMCLSTGLVLTFTROYASEINPHMEKLESEV 825
DB 2625 QYQELQLDUSE-----KVRAGVORLSVQSAISTQPEAVKQOLETESEIRSDQLDHEV 2678

RESULT 13
RA50_YEAST STANDARD; PRT; 1312 AA.
ID RA50_YEAST
AC P12753;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA repair protein RAD50 (153 kDa protein).
GN RAD50 OR YNL250W OR N0872.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=REB21;
RX MEDLINE=89276917; PubMed=2659437;
RA Alani E., Subbiah S., Kleckner N.;
RT "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT purine nucleotide-binding domain and two large heptad-repeat
RT regions."
RL Genetics 122:47-57(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeher U., Beinhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SU11
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae."
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: INVOLVED IN DNA DOUBLE-STRAND BREAK REPAIR (DSBR). THE
CC RAD50/MRE11 COMPLEX POSSESSES SINGLE-STRAND ENDONUCLEASE ACTIVITY
CC AND ATP-DEPENDENT DOUBLE-STRAND-SPECIFIC EXONUCLEASE ACTIVITY.
CC RAD50 PROVIDES AN ATP-DEPENDENT CONTROL OF MRE11 BY UNWINDING
CC AND/OR REPOSITIONING DNA ENDS INTO THE MRE11 ACTIVE SITE.
CC -!- SUBUNIT: FORMS A COMPLEX WITH MRE11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X14814; CAA32919.1; -
DR EMBL; X96722; CAA65494.1; -
DR EMBL; Z71526; CAA96157.1; -
DR FIR; S05808; BMBYDL.
```



```

Oy 409 RLILWFSHLSNFORIPWEENAFVLDLPK-----WAPKRVFQEIQLQREVRLSYWDK 461
Db 1310 RL-----TTDAET-----KVADLEKLOHEHDDHLIQGDLEKAL-KDSEKNFLRK 1356
Oy 462 IKOSTENATALEELLPPKAGPNFYSLBEGKEKTEBEOQLSARVREKQTA-----514
Db 1357 EAEMTEN-----IHSLEEGKEETKKE--IAELSSRLDNLQATNKLKNO 1398
Oy 515 -----RDMIVWIEETIYPVHGFEVLTIV--VOITLDDIGSKSFHLL 553
Db 1399 LDHLNQETRLKEDVLKESLIISLEESLNQRQKESLDDAKNLEHLDPTSRKNSSL 1458
Oy 554 VTLERYQGVSKLPDNDKQVMLSQVSTYKNNVQMTAVADRMMGYRLVSN-QAIVR 612
Db 1459 ---MEKIESINSL---DDKSELASAVE-----KIGALQKLHSELSLMEKNSQLO 1505
Oy 613 WYFSPENVDPQVHSQDPWEIIGN-----ALKNTYRISDLRKDISNTKNVLYVAEKSA 666
Db 1506 EAKERIQVDESTIOELDHEITASKNNYEGKLDKDSIIRLSENTEQL--NNLLAEKSA 1563
Oy 667 NARVELEAEESKL-----SLVEGEPIVLGENPAKMKRLKSTVEKTGEAELS 712
Db 1564 VKRLSTE-KESEILOFNRLADLEYHKQVSESE--LGRSKLK--LASTE---ELQLAE 1614
Oy 713 RESLEAKEALIN-----RALSETEVLLLLFOSFLGVKLERLPDPTKVRSVOD 760
Db 1615 NERLSLTRMLDLQNVKDLNKSLSLEDLRLSLSDSVASIOKECKIKSNTVESIOD 1674
Oy 761 -LKSTGA-----EDDKPSAMDVDSNGPNPKSCVEGEREQWCLSTIGVTAFTROYASEI 814
Db 1675 VLTSVQARNAELEDVSRSD-----KIRRRDRDCEHLGSKL-----1711
Oy 815 WPHMEKLESEVFGSDVHPLFQA 838
Db 1712 ----KKLHSQL----EEQHETFFRA 1728

RESULT 15
MYH3_HUMAN
ID MYH3_HUMAN STANDARD; PRT; 1940 AA.
AC P11055; Q15492;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic
DE myosin heavy chain) (SMRCE).
GN MYH3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=89263803; PubMed=2726495;
RA Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H.,
RA Rubinstein N.A., Kelly A.M., Sarkar S.;
RT "Nucleotide sequence of full length human embryonic myosin heavy
RT chain cDNA."
RL Nucleic Acids Res. 17:3591-3592(1989).
RN [2]
RP SEQUENCE OF 774-1940 FROM N.A.
RX MEDLINE=90033298; PubMed=2806346;
RA Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,
RA Raychowdhury M.K., Rubinstein N.A., Kelly A.M., Sarkar S.;
RT "Human embryonic myosin heavy chain cDNA. Interspecies sequence
RT conservation of the myosin rod, chromosomal locus and isoform
RT specific transcription of the gene."
RL FEBS Lett. 256:21-28(1989).
RN [3]
RP SEQUENCE OF 856-1940 FROM N.A.
RX MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,

```

```

RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
RT myosin heavy chains."
RL Eur. J. Biochem. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 856-1940 FROM N.A.
RX MEDLINE=89366648; PubMed=2771643;
RA Karsch-Mizrachi I., Travis M., Blau H., Weinand L.A.;
RT "Expression and DNA sequence analysis of a human embryonic skeletal
RT muscle myosin heavy chain gene."
RL Nucleic Acids Res. 17:6167-6179(1989).
CC -|- FUNCTION: MUSCLE CONTRACTION.
CC -|- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -|- DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
CC AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
CC MUSCLE.
CC -|- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -|- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -|- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -|- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).

```

FT CONFLICT 1608 1609 SR -> RA (IN REF. 3).
FT CONFLICT 1663 1664 RG -> QT (IN REF. 2).
SQ SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;

Search completed: August 20, 2002, 15:36:29
Job time: 597 sec

Query Match 2.9%; Score 128.5; DB 1; Length 1940;
Best Local Similarity 17.9%; Pred. No. 7.7;
Matches 163; Conservative 146; Mismatches 307; Indels 295; Gaps 41;

QY 15 GPEYGTSSDYKDHICTCFGVIREIERSGDQVLPFLQCAEOLPHKIPLYGTIGLNLN-- 72
DB 408 GNEYVTGQTVQVHHAVNALSKSVYEK--LFLMWVTRINQQLDTKLPRQ-HFIGVLDIA 464
QY 73 -----LENEDFVQKLVESVHANFQVALDSGNCN 100
DB 465 GFEIFEYNSLEQLCINFNTEKLOQFFNHMFVLEQEEYKKEGIEWTFIDF--GMDLAAC- 521
QY 101 SIRILLRFM--TSLCSKVTPASLIVFETLLSSAATTVDEE--KGNPSWQP----- 149
DB 522 -IELIEKPMGIFSILEECMFPRATDTSFKNL-----YDQHLGKSNFQPKPVKGRA 574
QY 150 QADFYVICILSSLPWGSSELAQVQDEIERVLGI-----QAYLSIRKNSSTSL 199
DB 575 EAHFSLIHAGTVDYSVSGWLEKNKQPLNETVTVGLYQKSSNRLLAHLATFATADADSG- 633
QY 200 NEFHNGEFESSLAEKD-----FVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLV 249
DB 634 -----KKVAKKKGSSFTQYSALFRENKLMSNLRT-----HPHF----- 670
QY 250 AGKFHELPIKM---EQSPSPS-DHSRAYSGKQKHDAL-----TRYPQRI-----R 292
DB 671 -----VRCIIPNETKTPGAMESHSLVHLQRCNGVLEGIRICRKGFPNRIYLGDFKQ 721
QY 293 RLNIIPANKMEDVQPTIDRFVVEYLLDVLFLYNGCRKECASYANLPVTF-RYEYLMAT 351
DB 722 RYVLNASALEGQFID-----SKRACEKLLASIDIDHTQYKFGHTKY 764
QY 352 LFSQIILLPOPPKTYLYTLVIM--DLCKALPGAFPAVAG-AVRALFEKISDLDMESRT 408
DB 765 FFKAGLL-----GTLEEMRDDRLAKLITRTQAVCRGFLMRVEFQK-----MVQRR 809
QY 409 RLILWFSHLSNPOFI--WPWEFAFVLDLPKWAQKVFQEILOREVRLSYWDKIKQSI 466
DB 810 ESIFCIQYINRSPMNVKHPWPKLFF-----KIKPLL 841
QY 467 ENATALEELLPPKAGPNFMYSLGEGKEKTEEQOLSALSRKVEKQOTARDMIVWIEIYI 526
DB 842 KSAETEKE-----MATKKEEFQKTDELAKSEAKRKELEEK----- 877
QY 527 PVHGFVTLTIIVVOTLLDIGSKSFTHLVTVLYRYGOVFSKLCPDNDKQVMLLSQVSTYWK 586
DB 878 -----LVTLVQE-----KNDLQLOVQAE-----S 896
QY 587 NNQVMTAVADRMMGYRLVSNQAIKRVFSPENVDQFHVSDQWEILGNALNKTYNRISD 646
DB 897 ENLLDAEERCQDLIAKAFQLEAKIKEVTERAEDEEEINAE-----LTAKKRKLEDCSE 950
QY 647 LRKDISNITKNVLVAEKASANARVELEAAESKL-SLVEGEFVLGENPAKMKRLKSTVEKT 705
DB 951 LKADIDDLLEUTLAKVEK-----EKHATENKVNKNTFEELSGLDETIAKTREKKALQ-- 1001
QY 706 GEAELSLRESLEAKALLNRLALSETEVILLILLFQSLFGLVKERLPDPTKVRVSQDLKSIG 765
DB 1002 -EAHQALDLDQAEEDKVN-SLANKTSKL---EQQVEDLESSLEQEKRLRYDLERNKRK 1055
QY 766 AEDDKPSAMD--VDSNGNPKKSCVEGERE-QMCL-----STLGYLTAFTROYASEI 814
DB 1056 LEGDLKLAQESILDLNDKQDLERLKKDKFYCOLQSKVVEQTLG--LQFQKK-IKEL 1112
QY 815 WPHWEXLESEV 825
DB 1113 QARIEELEEEI 1123

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 14:02:31 ; Search time 51.19 Seconds
(without alignments)
1591.790 Million cell updates/sec

Title: US-09-882-986-2

Perfect score: 4374

Sequence: 1 MSNWKTLRLRIGKEGPEYGT.....EDVHPLFLQAISALQFPLH 848

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3784	86.5	749	2 C84508	probable cap-bindin
2	813	18.6	790	2 S50082	nuclear cap bindin
3	636.5	14.6	740	2 T15197	hypothetical prote
4	475.5	10.9	780	2 T39057	hypothetical prote
5	279	6.4	858	2 A44919	GCR3 protein - yea
6	166	3.8	879	2 C71083	conserved hypothet
7	163	3.7	1413	2 T26467	hypothetical prote
8	145	3.3	1023	2 AE1643	ATP-dependent dsdn
9	139.5	3.2	3259	1 A56539	glantin - human
10	139	3.2	1388	2 T30335	hypothetical prote
11	138	3.2	1295	2 T24587	glantin - human
12	137.5	3.1	3225	2 T52300	hypothetical prote
13	137	3.1	1156	2 B70356	chromosome assembl
14	136	3.1	1133	2 T22976	hypothetical prote
15	135	3.1	4151	2 T13734	groovin gene prote
16	134.5	3.1	886	2 H69378	conserved hypothet
17	134.5	3.1	1818	1 S73852	hypothetical prote
18	133.5	3.1	1825	2 T42725	actin binding prot
19	133.5	3.1	1885	2 T30847	actin binding prot
20	133.5	3.1	2033	2 T30849	actin binding prot
21	133.5	3.1	2331	2 T25410	hypothetical prote
22	132	3.0	4063	2 T42993	probable spectrin
23	132	3.0	4101	2 T23630	hypothetical prote
24	131.5	3.0	1313	2 F96673	hypothetical prote
25	131.5	3.0	1864	2 F86378	protein F2139.12 [
26	131	3.0	1805	1 A64224	cag pathogenicity
27	131	3.0	1927	2 G64585	large tegument pro
28	131	3.0	2077	2 T43991	centromere protein
29	131	3.0	2663	1 S28261	

30	130.5	3.0	1252	2 A47213	beta-fodrin - huma
31	130.5	3.0	2364	1 A44159	spectrin beta-6 ch
32	130	3.0	1298	2 T24480	hypothetical prote
33	130	3.0	2077	2 T44178	large tegument pro
34	129.5	3.0	853	2 T51505	hypothetical prote
35	129.5	3.0	1312	1 BMBYDL	RAD50 protein - ye
36	129.5	3.0	1957	2 T38077	hypothetical coile
37	129	2.9	4131	2 T21085	hypothetical prote
38	128.5	2.9	1048	2 T19045	ras GTPase-activat
39	128.5	2.9	1207	2 T19041	ras GTPase-activat
40	128.5	2.9	1940	1 S04090	myosin heavy chain
41	128	2.9	1819	2 A71928	cag island protein
42	128	2.9	3660	1 S02041	dystrophin, muscle
43	128	2.9	3678	2 S28916	dystrophin - mouse
44	127.5	2.9	284	2 A60607	tropomyosin - fluk
45	127	2.9	964	2 G86491	hypothetical prote

ALIGNMENTS

RESULT 1

C84508

probable cap-binding protein [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: C84508

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallic

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: C84508

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-749 <STO>

A:Cross-references: GB:AE002093; NID:g4558659; PIDN:NA022677.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g13540

A:Map position: 2

Query Match 86.5%; Score 3784; DB 2; Length 749;

Best Local Similarity 99.1%; Pred. No. 1.le-229;

Matches 735; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MSNWKTLRLRIGKEGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLQCAEOLPHK 60

Db 1 MSNWKTLRLRIGKEGPEYGTSSDYKDHIETCFGVIRREIERSGDQVLPFLQCAEOLPHK 60

Qy 61 IPLYGLGLNLEDFVQKLIVESVHANFOALDSGNCNSIRILLRFMTSLCSKVIQ 120

Db 61 IPLYGLGLNLEDFVQKLIVESVHANFOALDSGNCNSIRILLRFMTSLCSKVIQ 120

Qy 121 ASLIVVFETLLSSAATTVEEKGNSWQPADFYICILSSLPWGSEAEQVPDEIERY 180

Db 121 ASLIVVFETLLSSAATTVEEKGNSWQPADFYICILSSLPWGSEAEQVPDEIERY 180

Qy 181 LVGIOAYLSIRKNSSTGLNPFHNGFEFSSLAEEKDFVEDLLDRIOSLASNGWKLESVPRP 240

Db 181 LVGIOAYLSIRKNSSTGLNPFHNGFEFSSLAEEKDFVEDLLDRIOSLASNGWKLESVPRP 240

Qy 241 HLSFPAQLVAGFHELRPIKMEQSPSPSDHSRAYSGKQKHDALTRYPORIRRLNFFPAN 300

Db 241 HLSFPAQLVAGFHELRPIKMEQSPSPSDHSRAYSGKQKHDALTRYPORIRRLNFFPAN 300

Qy 301 KMEDVQPIDRFVVEEYLLDVLFLYNGCRKECASYMANLPVTFRYEYLMAETLFSQILLP 360

Db 301 KMEDVQPIDRFVVEEYLLDVLFLYNGCRKECASYMANLPVTFRYEYLMAETLFSQILLP 360

Qy 361 QPPFKTYTTLVIMDLKALGAPFVAVAGVRALEKISDLDMRSRTLILWFSHLSN 420

Db 361 QPPFKTYTTLVIMDLKALGAPFVAVAGVRALEKISDLDMRSRTLILWFSHLSN 420

Db 361 OPPEKTYLTYVIMDLCKALPGAPPAVAVAGAVRALPEKISDLDMESRTRILLWFSHLSN 420
QY 421 FQIWPWEHAFVLDLPKAPKRVFQEIILQREVRLSYDWKIKQSIENATALELPPKA 480
Db 421 FQIWPWEHAFVLDLPKAPKRVFQEIILQREVRLSYDWKIKQSIENATALELPPKA 480
QY 481 GPNFMYSLBEGKEKTEEQQLSAELSRKVKQOTARDMIVWIEETIYPVHGFEVTLTIWQ 540
Db 481 GPNFMYSLBEGKEKTEEQQLSAELSRKVKQOTARDMIVWIEETIYPVHGFEVTLTIWQ 540
QY 541 TLLDICKSFTHLVTVLRYGVFSLQVFSKLCPCDNDKQVLLSQVSTYKNNVQMTAVADRMM 600
Db 541 TLLDICKSFTHLVTVLRYGVFSLQVFSKLCPCDNDKQVLLSQVSTYKNNVQMTAVADRMM 600
QY 601 GYRLVSNQAVIRVVFSPENVDQPHVSDQPEIILGNALNKTYNISDLRDKISNITKNVLV 660
Db 601 GYRLVSNQAVIRVVFSPENVDQPHVSDQPEIILGNALNKTYNISDLRDKISNITKNVLV 660
QY 661 AEKASANARVELEAAGSKLSLVEGEPVLGENPAKMKRLKSTVEKTGBAELSLRESLEAKE 720
Db 661 AEKASANARVELEAAGSKLSLVEGEPVLGENPAKMKRLKSTVEKTGBAELSLRESLEAKE 720
QY 721 ALLNRALSETEVLLLLLFQSF 742
Db 721 ALLNRALSETEVLYISLMQHM 742

RESULT 2
S50082
nuclear cap binding protein - human
C:Species: Homo sapiens (man)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 17-Nov-2000
C:Accession: S50082; A54748
R:Katoaka, N.; Ohno, M.; Kangawa, K.; Tokoro, Y.; Shimura, Y.
Nucleic Acids Res. 22, 3861-3865, 1994
A:Title: Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap binding protein
A:Reference number: S50082; MUID:95023141
A:Accession: S50082
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-790 <NAT>
A:Cross-references: GB:D33002; NID:gl865647; PIDN:BAA06769.1; PID:9577628
R:Zaurralde, E.; Lewis, J.; McGuigan, C.; Jankowska, M.; Darzynkiewicz, E.; Mattaj, J.W.
Cell 78, 657-668, 1994
A:Title: A nuclear cap binding protein complex involved in pre-mRNA splicing.
A:Reference number: A54748; MUID:94349369
A:Accession: A54748
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-751,'D',753-790 <I2A>
C:Genetics:
A:Gene: GDB:NCBP
A:Cross-references: GDB:393275; OMIM:600469
A:Map position: 9q22.3-9q22.3
C:Superfamily: human nuclear cap-binding protein
C:Keywords: nucleus; pre-mRNA splicing

Query Match 18.6%; Score 813; DB 2; Length 790;
Best Local Similarity 26.6%; Pred. No. 4.2e-43;
Matches 231; Conservative 174; Mismatches 326; Indels 136; Gaps 22;

QY 5 KTLILRTIGKEGPEYSSDYKDHETCFGVIIRREIERSGQVLPFLQCAEOLPHKTIPLY 64
Db 32 ESILCKVGEK-----SACSLNESLGLAGVLEADLPNYSKILRLCTVRLPEKTIY 86
QY 65 GTLIGLLINLENEFDVQKLVSIVHANFOALDGSNCNSIRILLRFMTSLICKSVKIOPASLI 124
Db 87 TTLVGLLNARNYNGFEVEMIRQLKESLKANNYEAIVYRFLSLVNVCHVIAAPSWV 146
QY 125 VFETLLSSAATTVDKEGNPSQPADFYVICILSLPWGSGELAQVDPDEIERYLVGI 184
Db 147 AMPENFVS-----VTQEDVYV--QVRDRWYVYAFSLSPWVGKELYEKDAEMDRIFANT 199

QY 185 QAYLSIRKNSSTSGLNFFHNGEPFESSLAEKDFVEDLLDRITQSLASNGKLESVPRPHLSF 244
Db 200 ESYLKRQKTHVPMQLQW---TADKPHQPEEYLDCLAQIQKLDKDRWOERHILRPLAF 256
QY 245 EAOLVAGKFHELAPKICMEQSPSPSDHSRAYSGKOKHDALTRYPORIRRLNIFPANKMED 304
Db 257 DSILCEALQNLPPF-----TPP-----PHTEDSVYPMRVRVFRMFEDYDDPGP 301
QY 305 VQP-----IDRFVVEEYLLDVLFLYNGCKEACASYMANLPVTFR--YEYLMAETLFSQILL 358
Db 302 VMPGSHSVERFVTEENLHCIIKSHWKRTCAAOLVSPGKNKIPNLNHIIVEVIFAEFLQ 361
QY 359 LPQPPKTYLTYVIMDLCKALPGAPPAVAVAGAVRALPEKISDLDMESRTRILLWFSHHL 418
Db 362 LPAPPHIDVMTTLLIELCKLPQSLPQVLAQATENIMYLRDWTNTTCVDFRINWFSHHL 421
QY 419 SNFQIWPWEHAFVLDLPKAPKRVFQEIILQREVRLSYDWKIKQSIENATALELPP 478
Db 422 SNFQFWSWEDNSDLSQDPSPKPKFVRELVKCMRLSYHORL-----DIVPP 471
QY 479 -----KAGPNFMYSL--EEGKETEEOQLSAELSRKVKQOTARDMIVWIEETYP----- 527
Db 472 TFSALCPANPTCIYKYGDESSNSLPGHSAVALCLAVAFKSKATNDEIFSILKDVDPNPQDD 531
QY 528 --VHGFE--VTLTIYVQTLIDICKSFTHLVTVLRYGVFSLQVFSKLCPCDNDKQVLLSQV 582
Db 532 DDDEGFSNPKLIEVFQVTLHLAAKSFHSFSAKAFHEVFKTLASDEGKLHVLRVMF 591
QY 583 TYMKNVQMTAVADIRMMGYRLVSNQAVIRVVFSP-----NVDQPHVSDQPEIILGNALNK 639
Db 592 EVRNHPQMTAVLDVKIRTIQIVDCAAVANNWIFSELSROFTRLFV---WEILHSTIRK 647
QY 640 TYNRISDLRKDISNITKNVLVAEKASANARVELEAAGSKLSLVEGEPVLGENPAKMKRLK 699
Db 648 -----MKNHVLKIQK-----ELEEAKEKLA-----ROHKRRS 674
QY 700 STVEKTGEAELS--LRESLEAKEALLNRALSETEVLLLLFQSLFGVLKERLPDPTKVRSV 758
Db 675 DDDRSSDRKDGVLVEEOIERLQEKVSAQSEQKNLFLVIFQRFINILTEHL----- 725
QY 759 QDLKSGAEDKPSAMDVDSDENGPNPKKSCVEGEREQCLSTGLYLTAFTRQYASEIWP 818
Db 726 -----VRCETDGTSLV-----TPWKNCIERLQOIFLQHHLIIQOYM 762
QY 819 EKLESEVFSGE-DVHPLFLQAISSALQ 844
Db 763 VTLENLIFTAELDPHILAVPQFCALQ 789

RESULT 3

T15197
hypothetical protein F37E3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 10-Dec-1999
C:Accession: T15197
R:Wamsley, P.; Bradshaw, H.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid F37E3.
A:Reference number: Z18306
A:Accession: T15197
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-740 <NAN>
A:Cross-references: EMBL:AF003132; NID:g2088676; PID:g2088677; PIDN:AAH54133.1; GSPI
A:Experimental source: strain Bristol N2; clone F37E3
C:Genetics:
A:Gene: CESP:F37E3.1
A:Map position: 1
A:Introns: 26/3; 198/3; 288/3; 363/1; 402/2; 428/3; 480/1; 548/2; 615/2; 662/3; 703/
C:Superfamily: human nuclear cap-binding protein

Query Match 14.6%; Score 636.5; DB 2; Length 740;
Best Local Similarity 24.4%; Pred. No. 4.7e-32;
Matches 182; Conservative 163; Mismatches 322; Indels 79; Gaps 20;

QY 11 IGKGPVGTSSDYKDHITCGVTRRTERSGDQVLPFLQCAEOLPHKIPLYGTGLGL 70
DB 21 IGKVGENTGSSIEC--NLDKLTAFHDDLEKRYASTIIDIAGCAIYLPNRVTVYTLVGL 78
QY 71 LNLNEDFVKLVSVHANFOVALDNGNCSIRILLRMTSLSCSVIOPASLIIVFETL 130
DB 79 LNSKNFNGGDDVEKLISEQDQLLSKQYQEAQNLAIFLDCDIGNSGVLTAAQSIGEYLSF 138
QY 131 LSSAATTVDEEKGNSWPOADFFYICILSSLPWGSSELAEQVDEIERVLVGIQAYLSI 190
DB 139 IAAAF-----ENMP--QVRNDYIQTVLRLCPWIGKELTEKAPOMENIGAICKYLEL 191
QY 191 RKNSTSGLNPFHNGEFSSLAEKDFVEDLLDRIQSLASNGWKLESVPRPHLSFEAQLVA 250
DB 192 RKNKHVALQVMREGSTQK--QEDYLESLSAQIEALRNADNVENHPIRHYSGFETLQD 249
QY 251 GKFEHLRPIKMEOPSPSDHSRAYSGKQKHDALTRYQRIIRL-----NIFPANKMED 304
DB 250 ALOHNLPSQSPHETS-----DMIIYPYPLVWFLRFQDADCSAFSSKPLPG 294
QY 305 VQIDRFVVEYLLDVL----PYLNGCRKECASYM---ANLPVTRYEYLVMAETLFSQIL 357
DB 295 DSSIDRFLEGEIAWIEKQFNKACARELLAFEAENPSVPIGF----LIFETIFGQML 350
QY 358 LLQPPFRTLYTILMDLCKALPGAFVAVAGAVRALFEKISDLDMSRTRLILWFSSH 417
DB 351 RLPHAPYPAIPHCSLVLELLKLPDDYQILVQVTECIYRRADSQPVCIIDRMVDFSFH 410
QY 418 LSNFOIIPWBEWAFVLDLPKAPRVFVQEILOREVRLSYWDKIKQSTENATALEELP 477
DB 411 LSNFOYRTYTDKCKLNDKADAFSGQIFVREVIERCRFRGSEYKIIAALPQ--DFVKIHP 468
QY 478 PKAGNFMYSLSEEGKEKTEEOQLSALSRKVKKEKQTARDMIVMI-----EETIYPVHGF 531
DB 469 --CSPEVRLIDE--EDTALVQRAETFTQFOEROPAEAFNLKSNDENDLPYNINEF 524
QY 532 EYTVIVVQTLIDIGSKSFTHLVTLVRLYGVQFSLCPDND--QVMLLSQVSTYKNNVQ 590
DB 525 ---GLFVVMVKMASKYSHNSFALFRVQTLTKVCDASELYQKLETLTYSCHKTNQ 580
QY 591 MTAVAIIDRMGVRVLSNQAIVRWVSPENVDOFHVSQDPW--EILGNALNKNTYNSIDL 648
DB 581 MLMILTDLKLLKQVDCSAVVGWLPD-EKMMQEH--DROWLFEVLNQALEKLTQINVVE 637
QY 649 KDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTVEKTGEA 708
DB 638 KDIKELTEK-----TENKIEEDDESIDKMEDE-----TKEEFKQDLEDLENN 683
QY 709 ELSLRESLEAKALLNRLALSETVELL 734
DB 684 KEKLERMTYFQKGLFNDFLIHAETLL 709

RESULT 4
T39057
hypothetical protein SPAC6G10.07 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
R:Accession: T39057
R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: 221824
A:Accession: T39057
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-780 <OLI>
A:Cross-references: EMBL:Z98603; NID:e1059110; PIDN:CAB11293.1; GSPDB:GN00066; SPDB:SPAC
A:Experimental source: strain 972h; cosmid c6G10
C:Genetics:

A:Gene: SPDB:SPAC6G10.07
A:Map position: 1
A:Introns: 84/2

Query Match 10.9%; Score 475.5; DB 2; Length 780;
Best Local Similarity 23.9%; Pred. No. 6.5e-22;
Matches 192; Conservative 149; Mismatches 299; Indels 163; Gaps 36;

QY 40 ERSGDQVLPFLQCAEQ-----LPHKIPLYGTGLLNLNEDFVKLVSVHANFOVALD 95
DB 67 EANDPETITLDCIQTTFIIPVKIPLATLIIRASLR-----VPLLEKAAAYFCIQYF 122
QY 96 SGNCSNI-----RILLRMTSLSC--SKVIOPASLIIVFEETLLSSAATTVDEEKGNSWOP 149
DB 123 T-NLNSFLYEAQRDLM--LICMSFALQPGTL-----RPLFSLLDADAISKETKPSW-- 172
QY 150 QADFVVICILSLPW-----GGSELAEQVDEIERVLVGIQAYLSIRKNSST--SGL 199
DB 173 -GDNFIILINLPYFIANNLDLGGKDFANEILDQCE-----IYVRHRKSSITLSNPL 224
QY 200 NFFHNGEFSSLAEKDFVEDLLDRIQSLA--SNGWKLESVPRPHLSFEAQLVAGKHELRLP 238
DB 225 SIHND-----LSEBEL--DLLYKQLILSRENDFTPPYISQPMKFFESDFV-----HIV-- 270
QY 259 IKCMEQPSPSDHSRAYSGKQKHDALTRYQRIIRLNIIPANKMEDVQPIDRFVVEYLL 318
DB 271 -----PVSPSPPEWTFQTPQONELPSFKRFFELFNFEIRTPDASDVAAASIFRDIV 324
QY 319 DVLFLYNGCRKECASYMANLPVTRY-----EYLVMAETLF 353
DB 325 DVINLEFNRVBAQAQVLTDLTDVYFYTKTALRGTPVNLPLNDPSSRKAEDIIVEAVL 384
QY 354 SOILLPOPPFTLYTILMDLCKALPGAFVAVAGAVRALFEKISDLDMSRTRLILW 413
DB 385 GELLGSQNTYKPVVYHSLIECCRIAPKILAPTGRVIRLMTSMSSDLPLQTLDRFDW 444
QY 414 FSHLSNQFIWPEWAFVLDLPKAPRVFVQEILOREVRLSYWDKIKQSTENATALE 473
DB 445 FSHLSNSNFHKNWNEIPDVELDLHPRKVFMRITITRELILSYSTRISDSLP-----E 499
QY 474 ELL---PPKAGNFMYSLSEEGKEKTEEOQL--SALSRKVKKEKQTARDMIVWEETIYP 527
DB 500 EURLLGEOPSGPNFYENETHPLVQSSQIIIEALRLKPLEE-----LDILLOSEE-- 551
QY 528 VHGFEVT-LTVIVVQTLIDIGSKSFTHLVTLVRLYGVQV---FSKLCPDNDKQVMLLSQVST 583
DB 552 IONSETSAVRLVNSCAYSILGSRFSFHALNVFEKHLNTLKHFSR--KSLDSEIEVDELFS 609
QY 584 YKNNVQMTAVAIIDRMGVRVLSNQAIVRWVSPENVDOFHVSQDPWEILGNALNKNTYNR 643
DB 610 FVKLQPFNAVMWLDKMLNYSIISITSIIEWLIK-QDVTIWSRS-YTWSL-----VNTTFNK 663
QY 644 I-SDLRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENPAKMKRLKSTV 702
DB 664 LAALRRSVSNKEDSSLINE-----ANEEKIEIVTNLL-----LSAURALI 703
QY 703 EKTGEAELSLRESLEAKALLNRLALSETVELLLLFQS-FLGVLKRLPDPPTKVRVSQDL 761
DB 704 SENAE-----NIWVSHWLNLMKLYVESNFSVSKKDTIEEANE--PVQEN 745
QY 762 KSIGAEEDKPSAMD-VD-----SEN 780
DB 746 TSEEQEDTKMQPVDADVDEQPSEN 768

RESULT 5
A44919
GCR3 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YM8564.07; protein YM9553.01; protein YM125W
C:Species: Saccharomyces cerevisiae
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 29-Oct-1999
C:Accession: A44919; S53055; S54494

Qy	95	DSGNCNIRILLRPMTSLTCSKVIQPASLIVPETILLSSAATTVDBEKGNPSPQOADPY	154
Db	137	ETGPWNKIKILRFLSIL--SPMFLVDELINLYKSFELSIELNNLDPGNN--VPLSEAI	192
Qy	155	VICITLSLPW-----CGSELAQOVDPDEIERVLVGIOAYLSIRKNSTSGLNFF--HNGE	206
Db	193	YTNLLNIPYLFPPNRNNDGLRTKVE-----LLAYVEQWLYVKTTDINLLREYGE	244
Qy	207	FESSLAE-----KDEVEDL-----	225
Db	245	PPYEMVELRVVLPNVKKA LINNLEQLNELFPDWNHLLTPQTGDEGFDALTLPSVDLKL	304
Qy	226	S---LASNGKLESV-PPHLSFEAQL--VAGFHELRPIKMEQSPSPDHSRAYSGKQ	279
Db	305	SFVALNKFQSVDSMMKTPRYAFHYVLPNSAGNFETVPIS-----TYAGOL	351
Qy	280	KHDALTRYPQIRRLNIFPANKMEDVQPIDREVVEVLLDVLFLYNGCRKE-----C	331
Db	352	FNDI-----IDLVESLENRKE-----VARQITLDLFFKAGITEPGESTAQLI	397
Qy	332	ASYMAN-LPVTFRYELMAETLSQIILLP--OPPFTLYTITVIMDLCKALPGAPVAV	388
Db	398	ATYEENLAPFTKIEDLAETILGLIFKLPSVSQPF-AFYITLLV-DICONSFKATAPVF	455
Qy	389	AGAVRALPEKISOLDMESRPTLLIWSHLSNFOFTWPEEWAFAVLDLKPWA-----PKR	443
Db	456	GRAFRFFYSHLDSLDPELKLURYLDWFISIQMSNFIKMKNEWED--DSIKFKGYFYNNKV	513
Qy	444	VFQVEILQREVRLSYWDKIKOSTENATALKEELPPK-----	479
Db	514	NFAKNLQKEQLRL-----SNFSESDSLQEFTKYLDTSYIPROOLINYOSLFT	564
Qy	480	-----AGPNFMYSLEEG--KEKTEBOQL-----SAELSRKYKKEQTARDMIVTEETI	525
Db	565	GYTVEEDSVRKNDLYFROEGVPHENTVRKILDYTHRANNSREYTELES-----ILGEL	617
Qy	526	YPVHGFEV-----LTIVVOTLTDGSKSFTH-----LVTVLERYGOVFSKCLCPD--NDKQ	574
Db	618	KNEYGSIISDFNRRVITLLVQAVTDSGSRSLSHANKYINDLKDLKTIIFAKTJELDETKE	677

[illegible]

Qy 674 AAESKLSLVE-----GEVLGNPAKMKRL-----KSTVEKTGBAEL-SLRSELEAKA 721
 Db 567 KAKKLESEIEDRLRLGFKTIDELSGRIRELEKFNHXYEAKNAEKELDILESLEKDERE 626
 Qy 722 LUNRAISE 729
 Db 627 ELDKAFEE 634

RESULT 7

T26467
 hypothetical protein Y11D7A.14 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
 C:Accession: T26467

R:Steward, C.
 submitted to the EMBL Data Library, October 1998

A:Reference number: Z20218

A:Accession: T26467

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-1413 <WLT>

A:Cross-references: EMBL:AL032632; PIDN:CAA21588.1; GSPDB:GN00022; CESP:Y11D7A.14

A:Experimental source: clone Y11D7A

C:Genetics:

A:Gene: CESP:Y11D7A.14

A:Map position: 4

A:Introns: 17/3; 62/3; 122/2; 178/3; 200/1; 312/1; 351/3; 387/1; 463/3; 561/1; 583/3; 62
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 F:35-696/Domain: myosin motor domain homology <NMO>

Query Match 3.7%; Score 163; DB 2; Length 1413;

Best Local Similarity 18.1%; Pred. No. 0.064;

Matches 160; Conservative 141; Mismatches 301; Indels 284; Gaps 39;

Qy 13 EKGPEYGTSSDYKDHICTCFGVIRREIERSGQVLPFL-LOCAEOLPHKIPLYGTGLGL 71
 Db 440 EGIKW-TQVNFANLQPTIDIEKPM-----GILSFLKEECV-VPN-----GSEKSL 486

Qy 72 -----NLEND-----FVQKLVSFVHANFOALDSGNCN---STIRLL 106

Db 487 EKLCSNLSDSFKSKOTQKCTIRHFYVQHVAGEVHYNDGWLKDRDNVETSLD 546

Qy 107 RMTSLCSKVQPA-----SLIVVFETLLSSAATTVDEEG 143

Db 547 SOSTHPLKLLPPVPVNNLTKRGITNTSVSYFLYKNOLCLDLTLNSSA----- 598

Qy 144 NPSWQPADFYVICILSSLPWGSSELAEOVPDEIERVLGIGQAYLSIRKNSSTGLNPFH 203

Db 599 -----HFIRCVSNY-----EKLPIKIDAPLVLAQ-----LKCNGVLEGIRICR 637

Qy 204 NGEFESSLAEKDFVEDLLDRIOQLASNGWKLESVPRPHLSFEAQLVAGKFHELPTKME 263

Db 638 EG-YPSRLSHSEFIE---RYSLLMKN----- 659

Qy 264 QPSPSDHSRAYSGQKHDAITRYQPIRRLNFIIPANKMEDVQPDIFVVEEYLLDVLY 323

Db 660 -----KEQSKGASEKEKCTLI-----CQDAQ----- 680

Qy 324 LNCGRKECASYMANLPVTFRYEVLMAET-LFSQIILLPOPFPKTLTYTLVIMDLCKALPG 382

Db 681 ---VRKE-----RYAVGKTKLFCXGVISELEYKRNRYI-----S 712

Qy 383 APPAVVAGAVRALFKISLDLDMESRTRILWFSHLSNFQFI--WPWEWAFVLD--LPK 438

Db 713 SPILIQANIRVLYNQKDLIERKKLEAVVTIQDNVQFAELSQMPWYIYHLTRGLIPR 772

Qy 439 WAPKRVFVQEIQREVLRSYWDKIKO-SIENATALELLPPKAGPNFMYSLGEEKTEE 497

Db 773 NRDKR-RIELENEKLEF--BEIQEMEIKNEEALKE-----NLKLSMLLDREKSEK 821

Qy 498 QOLSALSRRYKEKETARDMIWIEETIYPVHGFEVTLTYVQTLTLLDIGSKSFTHLTVYL 557

Db 822 VVQKELEEVKGR-----EKLLKEFERKTMEEQN-----EEIFNVL 863
 Qy 558 ER-YGVQFSKLPDND-----KQVMLLSQVSTYKNNVOMTAVADRM-MGYRLVSNQA 609
 Db 864 ERKYNQHKVKAMNDVLRREYERKIBOLNMEKTDLENENOKLRETNRDROSHYSNLEKEV 923
 Qy 610 IVRWVFSPEVNDQFH-VSDQPW-----ILGNALN-----KTYNRIISDLRKDISNIT 655
 Db 924 MEKSSLIDELQIOIKLSDENNEQRTIAKLETALDEDEKARFARQNTTIGOMKLSIELN 983
 Qy 656 KNVLVAEKASANAR-----VELEAAESKLSLVEGEPVLGENPAPKMRKLKSTVEKTGBAEL 710
 Db 984 EXIARFDNTALNERNSTRKIEREKEKLNELTTAKETIOQAQKIDELKEECCRKNKAS 1043
 Qy 711 SLRESLEAKALLNRALSETEVLLLLLQSFGLVKERLPDPPTKVRVSQDLKSGAEDDK 770
 Db 1044 RLERKLEDEAMMADCVKELK-----DSHKEKLE--MEQKVEDVKRKNK--- 1087
 Qy 771 PSAMDVDSENGNPKKSCVEGERE-----QWCLSLTGLVLTAFTRQYA 811
 Db 1088 -----LENENSTQKSOIETFORESSVDSYGRSSGRSLTLGROY 1128

RESULT 8

AE1643

ATP-dependent dsDNA exonuclease sbcC homolog sbcC [Imported] - Listeria innocua (st
 C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AE1643

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blc
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsi

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, J.

Ok, C.; Schlueter, T.; Simoes, N.; Tietre, A.; Vazquez-Boland, J.A.; Voss, H.; Wei

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AE1643

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1023 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC96917.1; PID:g16414173; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: sbcC

Query Match 3.3%; Score 145; DB 2; Length 1023;

Best Local Similarity 16.8%; Pred. No. 0.54;

Matches 155; Conservative 159; Mismatches 297; Indels 312; Gaps 38;

Qy 6 TLLLRIGEKGPEYGTSSDYKDHICTCFGVIRREIER---SGDQVLPFLQCAEOLPHKIP 62

Db 80 TLVFL-----KDKYQISRPQOEIAKORGNGTTTSPKAELEYELIGDEM 126

Qy 63 LVGTGLGLNLENDVQKLVSFVHANFOALDSGNCNSTRILLRF-----MTSLCSKVI 118

Db 127 LLASSVSDVNTKMEELIQL-----NVDQFRQILMPQGEFREILLVSDSK 170

Qy 119 OPASLIVVFTLLSSAATTVDEEK-GNPSWQPADFVICI-----LSSLPWGSSEL 169

Db 171 EK-----EVILORLAHTVYVEKVENLLWEKOKAEILVYEARKKVAELAEISLPQIEI 223

Qy 170 AEQVPDEI-----ERVLVIGQAVLS-IRKNSSTS-----GLNFF 202

Db 224 TCKTTTETISLQTEATKQEQAILAELESALSIRKETSEAVEKVTLAKEQLLDQNL 283

Qy 203 HNGFESSLAEKDFVEDLLDRIOQL--ASN-----GWKLESV--PRPHLSPEAQ 247

Db 284 IE-EVAKLEAEKDFYOVIANRIEAKRASNLSQDALCIRLKEQLETAFTAVTEKQVAHEVE 342

Qy 248 LVAGKHELRPIKCMQPPSPSDHSRAYSKQKHDAITRYQPIRRLNFIIPANK-----M 302

A;Residues: 1-1388 <BOL>
A;Cross-references: EMBL:X94082; NID:g1129172; PID:c213754; PIDN:CAA63826.1
C;Genetics:
A;Gene: klp2
C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

Query Match	3.2%;	Score 139;	DB 2;	Length 1388;	
Best Local Similarity	20.0%;	Pred. No. 2;			
Matches 174;	Conservative 156;	Mismatches 328;	Indels 214;	Gaps	
QY	35	IRREIERSGDVLPFLQCAEOLP----	HKTPLYGTILGLLNLENEDFVQKLVESVHAN	89	
DB	381	LQAEVKLKEQLSOLL--SGQMPGDISVARVPSVGD-----	NNMDYMNFTFAM----	427	
QY	90	FOALSDGNCNSIRILLRPMTSI--LCS----	KVIPASLIVVF-ETLLSSAATTVDDEKG	143	
DB	428	--MILEKSD-REKKVLQKVQVLEDLCKNKKFQISQNMIVKFPREDHISRLKKAHKRGRI	484		
QY	144	NPSWQPADFYVICLSSIPWGSSELAEOVPDEIRVLVGTQAYLSIRKNSSTSGLNFFH	203		
DB	485	SLSNNEQDDF-----IAELKEEIRTLKEQVHEHPRAKYALENHSLSREENKRLHSLSQSVK	539		
QY	204	NGE-----FESSLAEXDFEDLLDRITQSLASNGWKLES-----	VPRPH	241	
DB	540	RAQEVTAQIMIAELEKAFLEVSSEKD-----	RQVPMHSTPIQLDNNLSMSAARMERM	593	
QY	242	LSPEAOLVAGK-----FHRLRIKQMEQSPSPSDHSRAYSGKHDAHLTRYQPRIIR--LN	295		
DB	594	LQLESELATSKOYEYEFKELTKKKQVEQES-----	ELQSLIKSNQHLNILE	640	
QY	296	IFPANKMEDVQIDRFVWEYLLDVLFYNGCRKECASYMANLPVTR-----	YE	345	
DB	641	AIRANKRRESQSLNRHAAETIKNM-----	TTPTKSYNLSRLVPRLSPDAMPGLMD	693	
QY	346	YLMAETRFSQLLLPOPP-----	FKTLYYTL-VIMDLCKALPAGFAPVAVAGVRAALFEK	398	
DB	694	TPKSGDVMDIDINEPIPPMSQAEATAEELURIVQEQVTAQLAKLDEEGKNIR--IQQQ	752		
QY	399	ISOLDMESRTRLILWFSHLSNFQIWPWEW----	AFVLDLPKWAPKRVFVQETILOREVR	455	
DB	753	VNKLECS-TQIQELPNSERN--WNKEQODLQAIKSLEKQKQENKSOEDVLKSEVH	807		
QY	456	LSYWDIKOSIENATALEBLLPPKAGPNMYSLEGGKTEEQQLSA-----	502		
DB	808	-----DLRVVLOSAD--RELGAVGK-----	EYSLREKQEKELSOLSAHMDVQLQDNVR	856	
QY	503	-ELSRVKEKQOTARDIMVIEETIPVGHFEVTLIVVQTLIDIGSK-----	SPTH	553	
DB	857	LEHETLLEKRSLODAFDNLEE-----	VMKFEIQ--LKQEISD--SKHENETLRAEFSNL	908	
QY	554	VTVLERYGOVFSKLPDNDKQVYLLSQVSTYWKNNVQMTAVADRMGMRYLSVNOQATVRW	613		
DB	909	LELLE-----TEKERRKLTSLQEEDEKENKTKELLQVVDENMHLRKQCSSELATK-	957		
QY	614	VFSPENVQPHYSQDQWEILGNALNKNTYRISDLRK-----	DISN---ITKNVLV	660	
DB	958	--CEQQTVELHG-----	LEHSLSSKEMIADELEKKNTADKEVYVADLMNOIQVHRTTII	1008	
QY	661	AEKASANARY-ELEAAESKLSLV-----	EGEPVLGENPAKMKRLKSTVEKGT-----	EAE	709
DB	1009	HKTESIDLLTRELIEDHISKYSIVLAKESKTVIEEQEKQIEELRECLERKQSDADNIKE	1068		
QY	710	LSURESLEAKEAL--LNRALSETEVLLLLLFOSFLGVLKERLPDPT-KVRSVQDLKSGIGA	766		
DB	1069	LLCDDLAHATEELEKLEATEAFNKQEA--LLTHTEKELVEKEQOISELTNOVKLMTDLE-ISR	1126		
QY	767	EDDK--PSAMDVDS-----	ENGPNKS	786	
DB	1127	EQEKIRPASSNSSSPVLPETPRTEGPNPYS	1158		

RESULT 11

T24587
hypochemical protein T06E4.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24587
R:Lloyd, C.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19910
A:Accession: T24587
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1295 <WIL>
A:Cross-references: EMBL:Z70756; PIDN:CAA94789.1; GSPDB:GN00023; CESP:T06E4.1
A:Experimental source: clone T06E4
C:Genetics:
A:Gene: CESP:T06E4.1
A:Map position: 5
A:Introns: 17/3; 218/2; 385/3; 481/3; 946/3; 1034/3; 1222/2

Query Match	3.2a;	Score	138;	DB 2;	Length	1295;
Best Local Similarity	19;9a;	Pred.	No. 2.1;			
Matches	173;	Conservative	145;	Mismatches	307;	Indels 246; Gaps 42;
Qy	20	TSSDYKDHIEC	---	FGVIRREIBSRGDOVL	PFLQCAEQLP	PHKIPLYG-----TLIGL- 70
Db	206	TSQEKSRIEYSLEYEFMELEKNIEH	---DEVV-----	KDLQKIELG	EKHROT	LEIK 256
Qy	71	--LNLNEDEFVKVLVESVHANFOVALDSGNCN	SIRILLRFMTSL	CLCSKVIQ	PASL	VVFE 128
Db	257	EAREVEKMLMQQVEEA	V---KTKSBREA	KTLLEKRIE	LEMLFEP	NGEKIQFE 313
Qy	129	TLSSAATTVE	-----	EKNPMSQ	QADFVVIC	ILSSLPWGGSELA 170
Db	314	HQBELKSRCEBELDKALKVDVMOHS	IEDYKKFVELQ-----	-----	-----	EMK 354
Qy	171	EQVPEIERVLYGQA	---YLSIRKNS	TSGLNFFNG	FEFESSLA	BKDFVEDL-LDRIQ 225
Db	355	EEADEQLQKAKEDIETLQMKYVEL	---ETTINKVE	FSNSEIETL	KSHEIV	VRKMLMLDEIH 411
Qy	226	SLANGWKLESVPRPHLSFPAQLVAGF	HELURPTK	CHMQPSP	SD--HS	RAYSGKQK---- 280
Db	412	RLEN---EMSALQPKNDTTELEELQK	TLDDLLK-IDCCN	LTSKML	LOSEL	VEYVEKATSE 467
Qy	281	-HDALTYPQIRILNIFPANKMEDVQ	PIDRFVVEE	YLLDVL	FVLYNG	CRKECAS---YMA 336
Db	468	IGEAQVNGELLEQIN	---SURVENAK-----	---LVDMEG	OLDNAH	RAEDKDVRIIS 514
Qy	337	NLPVTFREYLMATFLFSQILLPOP	PKTYLYTLVTMD	CLKALP	GAPF	AVVAGAVRALF 396
Db	515	ELLTTI--ESLRQDSEASDKLLM	SESTQN-EYSLA	ENTVSEL-----	ETMREY	KASV 566
Qy	397	EKISDLMESRTRLILMFSHLS	FPQFIWPMEE	WAFVLDLP	KWAPK	RVFVQEILOREVRL 456
Db	567	DKVCSLQLEEE	---IQHETS-----	-----	-----	VELEEAERI 594
Qy	457	SYWDKIKOSINATALBELP	PKAGPMVSL	BSGKTE	BEQQLS	AEKSRKVKQKQFARD 516
Db	595	K-----ELELAQBEAV	--KTGSSQLK	LETIVQ	EDCQ-----	KLRDLQKEQI--Q 635
Qy	517	MIVWIEETIYPVHGFEVTLTIV	QTLIDIGSKS	FTHLVT	VLYERY	GOVFSKCLCPDNDKQVM 576
Db	636	QLVSLRETSEYMH	-----QESARH	OE---EKY-Q	OSKLM	STEAREVIE 674
Qy	577	LLSQVSTYWKNNQMTAVA	IDRMGMYL	SVNQAT	VRVFS	PENVDOFHV-----S 626
Db	675	LRSSID	-----	SLQAEVR	--VQSD	SAADQKHILEDYLRKTRQA 710
Qy	627	DQPEWILGNALNKTYNRIS	DURKO-----	ISNITKN	VLVAEK	SANARVLEEAESKLSV 682
Db	711	EETNEKLRSDLASBEEQ	ILDLUKNO	QOESLID	DKKELH	SASTNOELQVSEMLKIEVS-- 768
Qy	683	EGEPVLGNPAKMRKLKAST	VEKTG-EA-E	LSLRSE	BAKAL	LNRLASETEVLLLLLFQS 740

Db 769 -----NARQKVNESVLFKESFALQELASQOEVSRVVDAAVQEKDGLRL----- 815
Qy 741 FLGVLRKLPDPTKVRSDVODLKSIGAEEDDKPSAMVDSENGNPKKSCVEGE--REQWCLLS 798
Db 816 -VDTLKLKIEDTEK--SAQDLQSSVEETKQLQDLQ-----NPKQNAEVLSELNKLNSS 868
Qy 799 TLGYLTAFPTROYASEIWPHEKLESEVFSGE 829
Db 869 HKRDMVALASQ-----LEELQHKLVVGE 891

RESULT 12
152300
gi|152300|gb|U00000.1|genbank|gi|152300|gb|U00000.1|genbank|
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: 152300
R:Soeda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: 152300; MUID:95100974
A:Accession: 152300
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:9662389; PIDN:BA05025.1; PID:9808869
C:Superfamily: giantin

Query Match 3.1%; Score 137.5; DB 2; Length 3225;
Best Local Similarity 19.3%; Pred. No. 8.4;
Matches 168; Conservative 147; Mismatches 357; Indels 197; Gaps 32;

Qy 8 LLRIEKGPEYGTSSDYKDHITC-----FGVIRREIERSGDQV 46
Db 1427 LCENKQKPEEIGESRAKQIQKQLQAALISKEALKENKSLQEELSARGTIERLTKSL 1486
Qy 47 LPFLQCAEOLPKIPLVGLTGLLENEDFVKLVESVHANFQVALDSGNCNSIRILL 106
Db 1487 ADVESQVAQNKEKTVLGLRL-ALQOERDKLITMDRSLLENQSL---SSSCESLKLAL 1542
Qy 107 RFMTSLCKSVIQPASLIIVVFETLSS---AATTVDDEKGNPSWQPADFYVICILSSLPW 164
Db 1543 EGLTE-----DKEKLVKEISLAKSIAESTEENQEKHK---ELQKEVEILL----- 1585
Qy 165 GSELAEQVDPDEIRVLVGIQA-----YLSIRKNSSTGLNFHNGEFESSLAE-KD 215
Db 1586 ---QSYENVSNEARIQHVVAVROEKQELYGLKLRSTEA-----NKKETEKQLQEAQ 1635
Qy 216 FVEDLLDRIOSLA-SNGWKLESVPRPHLSFEAQLVAGKPFHELPI-----KMEQ--PSP 267
Db 1636 EMEEMKRNKRFAPKSKQKQKILEBENDRLRA-----EVHPAGDTAKECMETLLSN 1687
Qy 268 PSDHSTRAYSGKQKHDALTRYPQIRIRLNIFPANKMEDVQIDRFVVEEYLLDVLVFLNCG 327
Db 1688 ASMKEELERVMKEYETLSKKFQSLMS-----EKDSLSEVEVDLKHQIEG- 1731
Qy 328 RKECASYMANLPVTRYEYLMATLFSQILLPQPPFKTYTLTMDLCKALPGAFPAV 387
Db 1732 ---NVSKQANLEATEKHD-----NOTNVTEGTQSIQGE----- 1762
Qy 388 VAGAVRALFEKISDLDMESRTRLLWFSSHLSNFOFI----- 424
Db 1763 -----TEODSUSMSTRPTC-----SESVSAKSANP-----AVSKDFSSHDEINNYIQ 1806
Qy 448 ETLQREVLRSYWDKTK-----QSIEN-ATALELLPPKAGPNFMYLSEEGKEKTEEQ 499
Db 1807 QIDQLKRIAGLEEKQKNEKFSQILENKNWLLSQISTKQGLKMLQEVTKMNLNQ 1866
Qy 500 LSAELSRVKKEQKQARDMIVWIEETIY-----PVHGF-----EVLTIIVVQTLLDIGSKS 549
Db 549 -----EVLTIIVVQTLLDIGSKS 549

Db 1867 IQBELSRVTKLKATABEEKDDLEERLMNQLAELNGSIGNYCQDVTDAQIKNELLESEMKN 1926
Qy 550 FTHLVTVLERYGOVFSKLCIPDNDKQVMLLSQVSTYTKNNVQMTAVADIRMGYRLVSNQA 609
Db 1927 LKVCVSELEBEKQQLVK-----EKTKESEIRKEYLEKIQ-----GAQKEPQGNKSHAKEL 1976
Qy 610 IVRWFSPENVDQFHVSDOPWEITLGNALNKTYNRI-----SDLRKDISNITKNVLVAEKAS 665
Db 1977 QELLKEKQOEVKQLOKDCIRYQEKISALERTVKALEFPVQESQKDLITRENLAQAQVEHR 2036
Qy 666 ANARVELEAAESKLSLVEGEP--VLGENPAKMRKLKSTVEKTEGEAELSLSRESLAKALL 723
Db 2037 KKAQAEFLASFKVLLDDTQSEAARVLADNLKLKELQSNKE-----SVKSQMKQKQKDEDL 2089
Qy 724 NRALSETVLLLLLFQSLFGLVLERLPDPTKVRSDVODLKSIGAEEDDKPSAMVDSENGNP 783
Db 2090 ERLEQAEKHLKEKN---MOEKL-DALRREKVLHEETIG-----EQVTLN 2133
Qy 784 KKSCEVGEREQWCLSTGLVITAFTRQVYAS 812
Db 2134 KDKVEVQOQLOENLDSIVTQLAFTKSMSS 2162

RESULT 13
B70356
chromosome assembly protein homolog - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000
C:Accession: B70356
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: B70356
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-11156 <AOF>
A:Cross-references: GB:AE000699; NID:92983238; PIDN:AAC06839.1; PID:92983243; GB:AE000699
A:Experimental source: strain VF5
C:Genetics:
A:Gene: xcpC
C:Superfamily: chromosome segregation protein SMCI

Query Match 3.1%; Score 137; DB 2; Length 1156;
Best Local Similarity 21.0%; Pred. No. 2;
Matches 154; Conservative 105; Mismatches 244; Indels 232; Gaps 32;

Qy 218 EDLLDRIQSLASNGWKLESVPRPHLSFEAQ-----LVAGKFHELPIKCMQPSPPS 269
Db 391 EKLTEKLSLNKEKQBLE-IQRANLANKIERIKEDINKLISEREKEKEKEQ----- 444
Qy 270 DHSRAYSGKQKHDALTRYPQIRIRLNIFPANKMEDVQIDRFVVEEYLLDVLVFLNCG 329
Db 445 EIKRLKAIRKKEEELR--NLQELNIYEKRLSE-----VRKLEEVLEKGAIER 493
Qy 330 ECASYMANLPVTRYEYLMATLFSQILLPQPPFKTYTLTMDLCKALPGAFPAV-- 387
Db 494 EVRS-----FSD-----VSDVFKDIKGVYGSVE 517
Qy 388 -----VAGAVRALFEKISDLDMESRTRLLWFSSHLSNFOFI----- 424
Db 518 LIRVNPEHITAIVAGGRLKFIYVDEDEV-AKECIQAKRMNLGPFSPILNRRVVEE 576
Qy 425 -----WPWEWA--FVLDPKWPAPKRVFVQEILOREVLRSYWDK-IKQSIENATA----- 471
Db 577 RPLRPRTKGAVDFAVNLVEYDPK-----FEKVVKFVFGDTLIVENFESAIGIGNY 629
Qy 472 -----LEELLPPKAGP-----NFMYSLEE-GKEKTEEQQLSAELSRVKKEKQTA 514
Db 630 RMVTLEGELFEKSGVITGAVKPSGELNRYVEELQRLNAEEKLNEESIITQKIREI 689

Qy 515 RDMIVWIEETIPVHGFVTLTVVQTLIDIGSKSFTHLVTV-LBRYGVF----- 564
Db 690 RNLI--SEKT-----ALLKYSERKIEELSEGLEQYEEKFEKLENSKE 731
Qy 565 -----SKLCPNDKQVMMLSOVSTYWK--NNVQMTAVAIIDRMGMYRLVSNQAIYRWVFS 616
Db 732 YLKILEEKLINVEDKLABELAIEYEEKLNKLKEGDIKRH-----YS 776
Qy 617 PENVDQ-----FHVSDQWEI-----LGNALN-KTY-----NRI 644
Db 777 REGVEEKREYKVKQVSEIEXKSUNEIERELNKKTYELEYLEKIQEKREERYLTTRI 836
Qy 645 SDLRKDISNITKNVLVAEKASANARVELEAAESKL-----SLVEGEPVLGNP 692
Db 837 KSLKKEIENL---ILFKEKTLQ---EVKEAEVYDYIKOKEELEKILNLSKGLK 889
Qy 693 AKMRKLTVEKTGBAELSLESLEAKBALLNRAISETEVLLLLFQSLGVLKRLPDP 752
Db 890 IKEELK---EKIPEKENLK-VLEKIENTNEELKEYEDLKLGADEESIPKLKEKLKV 945
Qy 753 TKRSVODLKSIGAEEDDKPSAMDVDSNGNPKKSCVEGERQWCLSTLGLYLTAFTRQYAS 812
Db 946 TE--EIOKLGSVNFRPAEDYAEELKRFNDYKEKQOKLEESKAIKKLIETENKKRKVFL 1003
Qy 813 EIWPHMEKLESEVFS 827
Db 1004 EAFNOKNSLRIFS 1018
RESULT 14
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T22976; T23157
R:Lightning, J.
submitted to the EMBL Data Library, June 1994
A:Reference number: 219645
A:Accession: T22976
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <WIL>
A:Cross-references: EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2.6
A:Experimental source: clone F59A2
R:Burton, J.
submitted to the EMBL Data Library, October 1995
A:Reference number: 219700
A:Accession: T23157
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <W12>
A:Cross-references: EMBL:Z60514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A:Experimental source: clone K01A11
C:Genetics:
A:Gene: CESP:F59A2.6
A:Map position: 3
A:Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 710/3; 738/3; 795/2; 1008/3; 1

Query Match 3.1%; Score 136; DB 2; Length 1133;
Best Local Similarity 22.1%; Pred. No. 2.3;
Matches 102; Conservative 73; Mismatches 182; Indels 104; Gaps 21;
Qy 442 KRIVQVEI--LOREVRLSYWYDKIKOSIENATALELLPPK-----AGPNFMYSLLEG-- 491
Db 39 KNEVMQSLRSLQEQM-LEKCDALQAEVNEAKALREEIQAKYDDVTQKAERIQGLEESKK 97
Qy 492 -----KETEQQQLSAELSRVKVEKQATRDIV-----WIEETIPVHGF--EV 533
Db 98 VLESEKQAFNEKEQEREQEL-AKAMEKLNSEQNILDEVTKKLEQSEEEVLAARGAIQEL 156
Qy 534 T-----LTIVQTLIDIGSKSFTHLVTVLERGVQVFSKL-----C-PDNDKQVMML 578

Db 157 TEKLESEKETSTAKTELEAVSKKLDSSSETSLKERSDMIEAMKIQILINCEKQKDEAVELL 216
Qy 579 SOVSTYWKNNVQMTAVAIIDRMGMYRLVSNQAIYRWVFSFENVDQFHVSDQPWILGNALN 638
Db 217 KOKLEVEKN-----MSDVEVQKQLLLESTTSEMKOHAEEAIVKQLE 260
Qy 639 KTYNRISDLRKDISN--ITKNVLVAEKASANARV--ELEAAESKLSLVEGEPVLGNPAK 694
Db 261 EAQSISLENLKKDAERNLKTALDESDESAISEITQMEAAKKELEASEKEK--SELREQ 318
Qy 695 MKRLKST-----VEKTGEAELS-LRESLE-----AKEALLNRAISLSTEVLLLLLFQ 739
Db 319 MDRLOKVHNAQEDIQKLOKTWELEMAKIAKSTEDKELAREQLAGELNAKEDLKVVVEE 378
Qy 740 SFLG-----VLKERLP-DPTKVRSVODLKSIGAEEDDKPSAMDVDSNGN 782
Db 379 KHTGQIQAQALDDAEEKVKVLKEQLERAQSALESQELAS-SQKADKIOELEKELQNAQ 437
Qy 783 PKKSCVEGEREQWCLSTLGLYLTAFTRQYASEIWIWPHMEKLES 823
Db 438 KRSSELETANEMVRS-----LTATLENSNSETEILKQKLET 474
RESULT 15
T13734
groovin gene protein - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13734
R:Strumpf, D.; Volk, T.
J. Cell Biol. 143, 1259-1270, 1998
A:Title: Kakapo, a novel cytoskeletal-associated protein is essential for the rest
A:Reference number: 217746; MUID:99054752
A:Accession: T13734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4151 <STR>
A:Cross-references: EMBL:Y09430; NID:g4185883; PIDN:CAA70581.1; PID:g4185884
C:Genetics:
A:Gene: grv
A:Cross-references: FlyBase:FBgn0013733
A:Map position: 2
Query Match 3.1%; Score 135; DB 2; Length 4151;
Best Local Similarity 19.1%; Pred. No. 17;
Matches 183; Conservative 156; Mismatches 348; Indels 270; Gaps 43;
Qy 55 EOLPHKIPLYGTLGLLNLNEDFVQKLVSFVHAFQVALDSCNCH--SIRILLRMTSL 112
Db 2482 ERLLOSIAEHKPLLDKLNKTGTGALGALVADDGCAKINEILDTDNYAALRLERERQQA 2541
Qy 113 LCSKVIQASLTVVFEETLSSAATTVDEEKGPNWQPADYVICILSSLPWGSSELAEQ 172
Db 2542 LESALQESSQFSKLEGLMLRALANTVDQVN-----QLD----PLSALP---QKIREQ 2586
Qy 173 VPDE-----IERVL-----VGIOAYLSIRK-NSSTGSLNFPHNGEFES 209
Db 2587 IEDNDALMDDLKRDQAFSAVQRAANDVIKAGNADPAVDIKAKLEKLNLNWVQNA 2646
Qy 210 SLAEKDFVEDLLDRITQSLASNGWK-LESYPRPHLSFEAOLVAGKFKHELRIKMEQP-SP 267
Db 2647 TKKRGSLLDDIL-----SVAEPFQKLSNVMKTLKDLLEETL-----SCOEPFAAQ 2691
Qy 268 PSDHGRAYSGKO--KHDALTYPQRIRRRNIIEFANKMEDVQPIDREWVEYELDV----- 320
Db 2692 FODIKKQVALQEIHEIDQTKPE-VEQVRRHGSNLMMNMGEPDKPEVKKHIEDLNDWD 2750
Qy 321 -LFYLNCGKCKEASYSANLPTVTRYEYLMMAETLFSQILLPQPPFKTYTLTYIMDLCKA 379
Db 2751 NITALYAKREE-----NLIDAMEKAMEFHETLQNLKFLTKRAEDKFAHLGAVGSDI--- 2801
Qy 380 LPGAPVAVAGAVRALFEKISD-----LDMESRTRLLIWFSHHLSNFQFIPWEEW 430

```
Db 2802 -----DAVKRQIEQLKSFKDEVDPHMVEVEALNRQAVELTERTS-----PEQ 2843
QY 431 AFVLDLP-----KW-APKRVFVQEILQREVRLSYNDKIKQSI-----ENATALEEL 475
Db 2844 ASIREPLSVNRRWEALLRGVVEROKQLEHALLHGCQFOHALNELLVWINKTDTSTLDQL 2903
QY 476 LPPKAGP-----NFMYSLEE-GKEKTEEQQLSAELS-----RK 507
Db 2904 KPIPGDPOLLEVALKLVLANDIAHQNSVDTLNDAGROLIETEKGSVEASTTQEKLRK 2963
QY 508 VREK-----QTARDMIWTEETIYPVHGFEVTLTIIVQILLDI-----GSKSFTHL- 553
Db 2964 LNNEWKOLLQKASDRQHELEALREAHGYTAIEVQDILGWLGDYDAVIGASKPVGGLPETA 3023
QY 554 VTVLERYGOVFSKLCPDNDKOVMLLSQVSTYKNNVQM-----T 592
Db 3024 TEOLERFMEVYNELDENRKPVKETIQAGQGEYIKRQNMKVSSSNLQHTLTLTKQRWDVV 3083
QY 593 AVAIDRMGGRV-----SNQAIVRWVFSPEVNDQFHVSDQPWEILGNALNKTYNR 643
Db 3084 SRASDKKIKLEIALKEATEFHDTLQAFVWLTQAE-----KLLSNA--EPVSR 3129
QY 644 ISD-----LRKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGENP 692
Db 3130 VLETTQAQOMEHKVLOKDVSTHREAMLLDKKGTGTH---LKYFSQKQDVILIKNLLVSVQ 3185
QY 693 AKMKRLKS-TVEKTGEAELSLRESLEAKAL--LNRALSETEVLLLLLQF----- 739
Db 3186 HRWVRVSKAAERTRALDHGYKEAREFNDAWSGMMQYLQETEQLDQIIEEATASKEPOK 3245
QY 740 --SFLGVLKE-----RLPDTKVRSVQDLKSIGAEDDKPSAMDVDSENGNPKKSC 788
Db 3246 IKKYICKLKETHRLQLGAKQSVYDGT-MRTCKNLLERAPKGRD-VLD-----KMLIE 3295
QY 789 VGEREQWCLSTLGLYLTFTROYASEIWPHEKLESEVFGEDVHPLFLQAISSALQF 845
Db 3296 L--KEQM-----TRVWSKSIDRQKLEALLSGQ-----FSDALGELLDW 3334
```

Search completed: August 20, 2002, 15:27:37
Job time: 5106 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:58:56 ; Search time 32.69 Seconds
(without alignments)
633.616 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNWKTLRLRIGKPEYGT.....EDVHPLFLQAISALQFPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.5	3.0	2154	2	US-08-841-349-4
2	125	2.9	469	2	US-08-968-751-2
3	122	2.8	443	2	US-08-795-475-6
4	120	2.7	469	4	US-09-052-089A-1
5	120	2.7	893	4	US-09-019-160-2
6	119	2.7	1805	1	US-07-853-913-2
7	117.5	2.7	893	4	US-09-019-160-9
8	116.5	2.7	1786	4	US-08-973-462-8
9	116	2.7	708	4	US-09-019-160-5
10	116	2.7	893	4	US-09-019-160-6
11	116	2.7	893	4	US-09-019-160-7
12	116	2.7	893	4	US-09-019-160-8
13	113.5	2.6	3248	1	US-08-353-700-1
14	113.5	2.6	3248	5	PCT-US95-16216-1
15	112	2.6	1333	3	US-09-356-952-2
16	111.5	2.5	1312	2	US-08-592-126-148
17	111	2.5	893	2	US-08-706-702-3
18	111	2.5	893	3	US-08-706-706-3
19	110.5	2.5	2482	1	US-08-328-254-6
20	109.5	2.5	677	4	US-09-019-160-3
21	109	2.5	893	3	US-08-484-661A-2
22	109	2.5	893	3	US-08-656-864-2
23	109	2.5	893	5	PCT-US96-09641-2
24	108.5	2.5	911	3	US-09-356-952-6
25	108.5	2.5	1312	2	US-08-687-080-51
26	107.5	2.5	1388	4	US-09-572-191-2
27	105.5	2.4	955	1	US-08-006-676B-1

28	105.5	2.4	955	1	US-08-282-845-2	Sequence 2, Appli
29	105.5	2.4	955	2	US-08-428-414A-3	Sequence 3, Appli
30	105.5	2.4	955	5	PCT-US94-00324-1	Sequence 1, Appli
31	105.5	2.4	976	4	US-09-104-324B-4	Sequence 4, Appli
32	105.5	2.4	2089	1	US-08-418-893D-23	Sequence 23, Appli
33	105.5	2.4	2089	1	US-08-418-893D-24	Sequence 24, Appli
34	105	2.4	285	1	US-08-149-809-24	Sequence 24, Appli
35	104	2.4	753	4	US-08-026-143B-3	Sequence 3, Appli
36	104	2.4	753	5	PCT-US92-10621-3	Sequence 3, Appli
37	104	2.4	753	5	PCT-US94-02233-3	Sequence 3, Appli
38	103	2.4	753	1	US-08-712-241-3	Sequence 3, Appli
39	103	2.4	1711	3	US-08-369-822C-10	Sequence 10, Appli
40	103	2.4	1711	3	US-08-582-776C-10	Sequence 10, Appli
41	103	2.4	1711	3	US-08-434-831B-10	Sequence 10, Appli
42	103	2.4	2101	1	US-08-466-390-4	Sequence 4, Appli
43	103	2.4	2101	1	US-08-470-950-4	Sequence 4, Appli
44	103	2.4	2101	1	US-08-467-781-4	Sequence 4, Appli
45	103	2.4	2101	1	US-08-195-487-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-08-841-349-4
; Sequence 4, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO44700S0
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2154
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-841-349-4

Query Match 3.0%; Score 131.5; DB 2; Length 2154;
Best Local Similarity 17.9%; Pred No. 0.0061;
Matches 179; Conservative 151; Mismatches 283; Indels 389; Gaps 47;

Qy	118	IQPASIYVVFETLLSSAATTVDKGNPSWQPADFYVICILSSLPWGSELAEQVPEI	177
Db	285	IETEKMIKYEYETLASDLEWIEQ-----TIIILNNRKFANS-----	320
Qy	178	ERVLVGIQAYLSIRKNSSTSGLNFFHNGEFSSLAEKDFVEDLLDRIQS-LASNGWKLES	236
Db	321	---LVGVOQQL-----QAFNTYRTVEKPKFTEKGNLEVLLEFAIQSKMRANNOKV-Y	368
Qy	237	VPRPHLSFEAOLVA-----GKHELRPIKMEQSPSPSDHSRAY---	275
Db	369	MPR-----EGKLISDINKAWERLEKAEHERELALNELIRQEKLOARFORKAAAMRET	423
Qy	276	-----SGKQKHDA-----LTRYPQIRIRRLNIFPANKME-----D	304
Db	424	WLSENRILVSQDNFGFDLPAAVEATKKHEATETDIAAEYERVOAV-VAVARELEAENYHD	482
Qy	305	VQPI-----DRV-VVEYLLDVLFLYNGCKEACASYANLPVTFREYLMATLTF-----	353
Db	483	IKRTARNDNVIRLWEYLLEL-----RARRQ-----RLEMLGLQKIFQEMLYIMDWM	532
Qy	354	-SQILLPQPFKTYLYLVIMDLCK--ALGAPFAVAVAGAVRAL-----	395
Db	533	EMKVLILLSQDYGKHL---LGVEDLLQKHALVEADIAIQAERVGVNANASAKQFATDGEYK	589
Qy	396	-----FEKISDLDMSRTRL-----ILWFSHLSNFQFIWPEE---WA	431
Db	590	PCDQVIRDRVAHMEFCVQELCOLAARRARLESRLW-----KEFWMAEEGW-	640


```
QY 432 FVLDPKWPKRVRVQEILOREVRLSYMD-----KIKOSIENATALEELLPPKAGNPFM 485
DB 641 -----TREKEKILSSDDYGDGLTSVWRLSKHRAFEDEMSGRSG-HFE 682
QY 486 YSLEEGK-----EKTEE-----QOLSAELSRKVK-----KOTARD 516
DB 683 OAIKEGEDWIAEBHFGSEKIRIRIIRQEWANLQSLSAIRKKRLEASLLHQFOADADD 742
QY 517 MIVWTEETIYPVHGFEV-----TUTIV-----VOTLLDIGS----- 547
DB 743 IDANMLDLIKVSSNDVGHDEYSTOLVKYKHKDVAEEITNCRPTDITLHEQASALPQAH 802
QY 548 -----KSFTHLVTVLERYGQ---VFSLCPDNDQVWLLSQVSTYWKNN 588
DB 803 ESPDVKGRLAGIEERCKEWAELTRUKQALQDTLALYKMFSEADACELWIDE-KEOWLNN 861
QY 589 VOM-----TAVADRMMGYRLVSN-----QAI 610
DB 862 MOIPEKLEDLEVIQHRFESLEPEMNNQASRVAVVNOIARQLMHNGHPSEKEITRAQODKLN 921
QY 611 VRW-----VSPENVDQFHVSDQWELGNALNKTYNRISDLRK-----DI 651
DB 922 TRWSQFRELVDKRDALLSALSQIYHLE-----CNETKSCIREKTKVIBSTQDL 971
QY 652 SNITKNVLVAEKASANARVELEAESKLSLV--EGEPVLGENPAKMKRLKSTVEKTGEAE 709
DB 972 GNDLAGVVALOCKTGMERDLVAIEAKLSDLQKEAEKIESEHPDQQAAILSLAISDVW 1031
QY 710 LSLRESLEAKALLNRAISETEVLLLLL--FOSFLGVILKERLPDPTKVRVQDLKSIGAE 767
DB 1032 BEMTKTKNREASLGEA-SKLQQFLRDLDDFOSWLSRTQTAI-----AS 1074
QY 768 DDKPSAM-----DVDSNGNPKKCEVGEREOWCLSTLGLTA---FTRQ 809
DB 1075 EDMPTNLTAEKLLTOHENIKNEIDNYEDYQKMRDMGE-----MYTQGTDAQYMFRLQ 1129
QY 810 YASET---WPMHEKL---ESEVFGEDVHPLFOAISALQF 845
DB 1130 RLQALDTGWNELHKWENRQNLSSOHAYQQFLRDTKQAEAF 1171

RESULT 2
US-08-968-751-2
; Sequence 2, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfield, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,751
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Giotta, Gregory
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX1024 GG
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 262-8710
; TELEFAX: (510) 222-9758
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 469 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-968-751-2
```

Query Match 2.9%; Score 125; DB 2; Length 469;

Best Local Similarity 21.6%; Pred. No. 0.0018;

Matches 90; Conservative 60; Mismatches 180; Indels 86; Gaps 14;

QY 392 VRALFEKISDLMESRTRILMFSH--HLSNFQFTWPEWAFVLDLPKWPKRVRVQ-- 447

DB 3 IRALCTICSDFFDHSRDVAAIHCGHFTFL---QCLIQWFE-----TAPSRCTPCQR 50

QY 448 -EILQREVRLSYMDKIKOSIENATALEELLPPKAGNPFMYSLSEEGKEKTEEQOLSAELSR 506

DB 51 IQVGARTINKUFFDLAQEEENVLDAEFL-----KNELDNVRALQSLQ 92

QY 507 KYKERQTARDMIVWIEETIYPVHGFEVTLTVVQTLDDIGSKSFTHLVTVLERYGVFSK 566

DB 93 KDKERDSQ-----VIIDTLRDTLEERNATVVSLLQALGKA-EM 130

QY 567 LCPDNDKQVWLLSQVSTYWKNNVQMTAVADRMM-----MGYRLVSNQAIYRWVFSFENV 621

DB 131 LCSTLKKQMKYLEQOQDETQKQEEARRLSKMKMTMEQIELLOSQRPEVEEIMRDMGV 190

QY 622 QHVSQDQWELIGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAESKLSL 681

DB 191 QSAV--EQALVTCVSLKKEYENLKEARKASGE-----VADK----LRKDLFSRSKQLQ 238

QY 682 VEGEVLGENPAKMKRLKSTVEKTGEAEALSRESLEAKALLNRAISETEVLLLLLQSF 741

DB 239 VYSE--LDQAKLELKSQADLQSAQKEIWSLKKKLTMLQETLNLPPVASETVDRVLVS- 295

QY 742 LGVLKERLPDPTKVRVQDLKSIGAEADDKPSAMDVDSENGNPKKCEVGEREOWCL 797

DB 296 -----PAPVEVNLKLRPFRDDIDLNFATFDVDTPPARPSS--OHGYIEKLCL 342

RESULT 3

US-08-795-475-6

; Sequence 6, Application US/08795475

; Patent No. 5965390

; GENERAL INFORMATION:

; APPLICANT: Bjivrock, Lars

; APPLICANT: Sjivbring, Ulf

; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/795,475

; FILING DATE: 11-FEB-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mcmasters, David D.

; REGISTRATION NUMBER: 33,963

```

; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-795-475-6

Query Match      2.8%; Score 122; DB 2; Length 443;
Best Local Similarity 20.0%; Pred. No. 0.0032;
Matches 90; Conservative 83; Mismatches 167; Indels 110; Gaps 20;

QY 452 REVLRSWDK-IKOSIENATALELLPPKAGPNFMS--LEEKGKTEEQQLSAELSRKV 508
Db 21 QNIRLRHENDLKLARLENAMEV-----AGRDFKRAEELEKAKQALEDQR--KDLETKL 71
QY 509 KEKOTARDMIWIEETIYPVHGFEVTLIVVQTL---LDIGSKSFTHLVTVLERYGQVFS 565
Db 72 KELOODYDLAK--ESTSWDRQRLEKEEKEKEALELAIDQASRDY--HRATALE----- 121
QY 566 KLCPDNDKQVNL-LSOYS-TYKKNV--QMTAVAIIDRMGMGYRLVSNQAIVRWFVSPEN- 619
Db 122 KELEKKKALELAIDQASQDQNRANRVLEKELETITRQOEINRNLLGNAKLELDQLSSEKE 181
QY 620 -----VDQFVSDQPEILGNALNKTYNRISDRKQISNIT-----KNVLVAEK 663
Db 182 QLTIEKAKLEEKQISDASRQSLRDLDSAREAKKQVEKDLANTAEKDKVKEDQISDA 241
QY 664 ASANARVELEAAE-----SKLSLVEGEVPLGENPAKMKRLKSTVEKTGEAE 709
Db 242 SRQRLRDLDSAREAKKQVEKDLANTAEKDKVKEDQISD--ASRQRLRDLDSAREAK 299
QY 710 LSLRESLE-----AKEALLNRALSETVL-----LLLLFQSLGVLKERLPDPKVR 756
Db 300 KQVEALEERANSKLALEKLNKELESKKLTKTEKAEKALQAEKALKEQL-----AK 354
QY 757 SVQDLKSIGA-----EDDKPSAMVDVSENGPNKPSCEGEGEQWCLSTGLYLTAFTR 808
Db 355 QAEELAKLRAGKASDSQTPDKPNKAVPGKGQAPAGTRPNQKA-----400
QY 809 QYASEIHPHMEKLESEVFGSDVHPLFLQA 838
Db 401 -----PMKETKRLPSTGETANPFPTAA 423

RESULT 4
US-09-052-089A-1
; Sequence 1, Application US/09052089A
; Patent No. 6346605
; GENERAL INFORMATION:
; APPLICANT: Lee, Soo Y.
; TITLE OF INVENTION: SIGNAL TRANSDUCER FOR THE TNF RECEPTOR SUPER
; FAMILY, AND USES THEREOF
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052.089A
; FILING DATE: 31-Mar-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26.742
; REFERENCE/DOCKET NUMBER: 600-1-198 CIP 1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-3600
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 469 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-052-089A-1

Query Match      2.7%; Score 120; DB 4; Length 469;
Best Local Similarity 20.7%; Pred. No. 0.0055;
Matches 87; Conservative 65; Mismatches 174; Indels 94; Gaps 14;

QY 392 VRALFEKISDLDMESRTPLLWFSH--HLSNFQFIWPEWNAFVLDLPKWPKRFFVQ-- 447
Db 3 IRAITCTCSDFDHSRQVAAIHCGHTFFHQ-----CLIQSFETAPSRCTCPQR 50
QY 448 -EILQREVRLSYWKIKQSIENATALELLPPKAGPNFMSLEEKGKTEEQQLSAELSR 506
Db 51 IQVGKRIITIKLFFDLAQEENVDREPL-----KNELDNVRAQLSQ 92
QY 507 KVKEKQIARDMIVWIEETIYPVHGFEVTLIVVQTLIDIGSKSFTHLVTVLERYGQVFSK 566
Db 93 KDKEKRDSQ-----VIIDTLROTLEARNATVVSLOQALGKA-EM 130
QY 567 LCPDNDKQVMLLSQVSTYWK-----NNVQMTAVAIIDRMGMGYRLVSNQAIVRWFVSP 617
Db 131 LCSTLKKQMKYLEQQQDDTKQAEAGRLRSMKMTMEIQLLQSQLEPEVEEMIR----D 186
QY 618 ENVDQFHVSDQPEILGNALNKTYNRISDRKQISNITKNVLVAEKASANARVELEAAES 677
Db 187 MGVCQSAV--EQLAVYCVSLAKEYENLEAKASGE-----VADK-----LRKDLFESSR 234
QY 678 KLSLVEGEVPLGENPAKMKRLKSTVEKTGEAEALSRESLEAKEALLNRALSETVLLLL 737
Db 235 KLTQVYSE--LDQAKLELKSQKDLQSDAKKEIMSKKLTMLQETLNLPPVASETVDRLV 292
QY 738 PQSLGLVKERLPDPKTVRSVQDLKSCAEDDKPSAMDVDSNGPNKPSCEGEGEQWCL 797
Db 293 LES-----PAPVEYNLKLRRPFRDDIDLNATFDVTPPPARPSS--QHGYIEKLCL 342

RESULT 5
US-09-019-160-2
; Sequence 2, Application US/09019160
; Patent No. 6306588
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; APPLICANT: Solus, Joseph
; APPLICANT: Yang, Shuwei
; TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic
; Nucleic Acid Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
; STREET: 1100 New York Ave., N.W., Suite 600

```

```

; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,160
; FILING DATE: 06-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: (To be assigned)
; FILING DATE: 06-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/037,393
; FILING DATE: 07-FEB-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.4250002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 893 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-09-019-160-2

Query Match          2.7%; Score 120; DB 4; Length 893;
Best Local Similarity 18.6%; Pred. No. 0.017;
Matches 99; Conservative 92; Mismatches 188; Indels 154; Gaps 22;

QY  422 QFTWPEEWAFLDLPKWAQRVVFQELQREVRLSY-----WDKIK--- 463
      | : : : : | : | : | : | : | : | : | : | : | : | : |
Db  206 QLGLGRNLEYILEHARELPPQRVKALLRDREVAILSKKLATLVNAPVEVDWEEMKYRG 265
      | : : : : | : | : | : | : | : | : | : | : | : | : |
QY  464 -----QSIENATALEBLPPKAGPNFMYSLEEKGTEKEOQLSAELSRKVKEQ 512
      : : | : : | : | : | : | : | : | : | : | : | : | : |
Db  266 YDKKKLLPILKEFAISIMKELQLYEAEP-----GYEIVDKHTFDLIEKLKEVP 318
      : : | : : | : | : | : | : | : | : | : | : | : | : |
QY  513 T-ARM-----IWMETITYPVHGFEVL-----TIYVTLLDIGSKSF 550
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db  319 SFALDLETSSLDPCNCEIGVISGFKPKTAYIPLHHRNAQNLDLTLSLKLEIDPS 378
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY  551 THLVTVLERGVQFSKLCPDNDKOVMLLSQVSTWKN-NVOMTAVID----- 597
      : : | : | : | : | : | : | : | : | : | : | : | : |
Db  379 SKIVGQNLKY-----DYKVLAVKGISPYPHFDTMIAAYLLEPNKKFNLEDLS 427
      : : | : | : | : | : | : | : | : | : | : | : | : |
QY  598 -RMWGRLVSNQAIVRWV-----ESPENVDFHVSDOPWEILGNALN 638
      : : | : | : | : | : | : | : | : | : | : | : | : |
Db  428 LKFLGKNTSQEQLMSFPSPFGSFDADVPVDKAANYCEDAD---ITYRKLISMKLH 484
      : : | : | : | : | : | : | : | : | : | : | : | : |
QY  639 K-----TYNRISD-----LRKDTSINTKTNVLAEKASARAVELEAAESKSLVEGEVP 687
      : : | : | : | : | : | : | : | : | : | : | : | : |
Db  485 EAELENVFIREFMPLVNVLARMELNGVVYDFLKKLSEEGKKEELEAKIYQTAGEPF 544
      : : | : | : | : | : | : | : | : | : | : | : | : |
QY  688 LGENPAKMRL-----KSTVEKTEGAELSLESLEAKEALNALSLETEVLLLLLF 738
      : : | : | : | : | : | : | : | : | : | : | : | : |
Db  545 NINSPKQVSILPEKLGIKPRGKTKTG--EYSTR--IEVLEEIAN---EHEIVPLILE 596
      : : | : | : | : | : | : | : | : | : | : | : | : |
QY  739 QSFLGVLUKE-----RLPDP TKVRSVQDLKSIGAEDDKPSAMD-----VDSENGPK 784
      : : | : | : | : | : | : | : | : | : | : | : | : |
Db  597 YRIQIKLKSYTDIFPLKLVNPKTRIHAHSFOTGTATGRLSSSPNQLNLTPTKSGEGEI 656

```

```

QY 785 KSCVEGREQCLSTLGYLTAFTROYASEINPHM---EKLESEVSGEDVHPL 834
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 657 RKAIVPODPDWIVSADY-----SOIELRIHLHSGDENLVKFAFEGIDVHTL 704

RESULT 6
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: Mckay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-853-913-2

```

```
Query Match      2.7%; Score 119; DB 1; Length 1805;
Best Local Similarity 20.8%; Pred. No. 0.077;
Matches 150; Conservative 110; Mismatches 241; Indels 220; Gaps 37;
```

Qy	218	EDLDRIQSASNGW--KLES-VPRPHLSFQAQLVAGKFHELRIPIK-----CMEQPS	266
	:	: : : : : : : : : : :	:
Dd	57	DDELASRLIVDQRWEKLEAEVQRDNLAEEVSAGRCQQVLARENTVQEACSRRAL	116
	:	: : : : : : : : : :	:
Qy	267	PPSDHSTRAYSQKHDAITRYQPQIRLNIPFAPKNMEDVPIDRFVVEYLLDVLFYUNG	326
	:	: : : : : : : : : : : :	:
Dd	117	EAKENARGWLSTQAAE-LEREALR-----AHEEERA-----HLN-	152
	:	: : : : : : : : : : : :	:
Qy	327	CRKSCASYMANLPVTFRYEYLMAETLFSQILLPFPFKTYLYTLVMIDLCKALPGAPPA	386
	:	: : : : : : : : : : : :	:

Db 153 AQAACAPRRP-----PAPPHR-----IPGPAPE 175
Qy 387 WV-----AGAVRALFEKISDLM---ESRTRLLWFSSHLSNFQIWPWHEWAFV 433
Db 176 VEDLARLGEVWRGAVRDYQERVAHMSSLGQAREL-----SQVARGARECR-- 223
Qy 434 LDLPKWAPKRVFVQ---EILQREVRLSYWIKOSIENATALEELLPPKAG--PNFMYSL 488
Db 224 LEVQOLQADRLSLOERREALBORLEGRWODRLQATDRQLAVEALEQKQGLSQIOAIL 283
Qy 489 EGEKE-----KTEEQQSIAELSR-----KVK-----EKQ 512
Db 284 EGGQOLAHKMSLSLEVATYRTLEAENSRLQTPGRGSOASGLGLDPLKPNFLGIPEDQ 343
Qy 513 TARDMIVWIEETIYP---VHGFEVTLTIVVOTLLDIGSKSETHLVTVLERYGQVFSKLP 569
Db 344 YLGSVLPALSPSPPLNLTPTVTAFLKTQBFLOARTPTLASTPIP---PISEAPCP 400
Qy 570 DNOK---QVMLLSOVSTYWKNNVQMTAVAIDRMGMGYRLVSNQAIWRWFSPENVDQPHVS 626
Db 401 PNAEVRAQEVPLSLLOQAPEPLWKAT-----VPSSAILPELEPPGGKQGHFP 451
Qy 627 DQWPEILGNALNKTYNRISDLRKDISNITKNVLVAEKASANARVELEAAESKLSLV----- 682
Db 452 D-----DLTSLATN-LNPHHTLEAK-DGESSESRVSSIFOED 487
Qy 683 EG---EPVLGENPAKMKRLKSTVKTGTGAELSLRESLEA-----KEALLNRLALSETEVLL 734
Db 488 EGOIWEIVKEADIEVKNSSAOKTQESGLDTEETQDSQGLQKETL--KALGE-EPLM 544
Qy 735 LLLFQSLGLVKKER-----LPDPTKVRSVQDLKLSIGAEDDKPSAMDVDSENGNP 783
Db 545 SLKIQNYETACKENCNCSSTEGHLGTLEGPEKEQI-PLKSL---EKNVSEKTLNCGVP 600
Qy 784 KKCEVGE-----REOMCLSTGLYLTFT---ROYASEINPHME-KLES-EVFSGEDVHP 833
Db 601 VLSLCKEDTRTEDQELMSPKGLTKRFSSLGKESQEVVRPSKEGNLESWTAKEESQHP 660
Qy 834 L 834
Db 661 L 661

RESULT 7
US-09-019-160-9
; Sequence 9, Application US/09019160
; Patent No. 6306588
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; APPLICANT: Solus, Joseph
; APPLICANT: Yang, Shuwei
; TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic
; TITLE OF INVENTION: Nucleic Acid Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,160
; FILING DATE: 06-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: (To be assigned)
; FILING DATE: 06-JAN-1998

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/037,393
FILING DATE: 07-FEB-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942.4250002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 893 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-019-160-9

Query Match 2.7%; Score 117.5; DB 4; Length 893;
Best Local Similarity 18.4%; Pred. No. 0-031;
Matches 132; Conservative 117; Mismatches 244; Indels 225; Gaps 31;

Qy 298 PANKMEDY-QPIDRVVVEEYLDVLFYLNCGCRKECASYMANLPVTFRYEYLM---AETLF 353
Db 31 PTNAVYDARMVLKFIKEHIIPEKDYAFAVAFOKAA-----TFRHKLVSCKAORPK 82
Qy 354 SQIILLPOPPF-KTLYTYLIVMDLCKALPG--AFPVAVAGAVRA----- 394
Db 83 TPALLVQOLPYIKRRIEALGFKVL--ELEGYEADDIATLAVRAARFLMRFSLITGDKDM 140
Qy 395 -----LPEKISDLDMESRTR-----LILFSSHLSNP----- 421
Db 141 LQVNEKIKVNRIVKGISDLELYDSKVKERYGVPHQIPDOLLALTGDDIDNIPGVTGIG 200
Qy 422 -----QFTWPEEWAFLVDLPKNAKRVFQVQILOREVRLSY-----WDK 461
Db 201 EKTAVQLLGKYNRLYEIHAHELPOVRKALLRDREAVILSKKLATLVNAPVEVDWEE 260
Qy 462 IK-----QSTENATALEEL-LPPKAGNFMYSLEEGKTEEBEQOOLSASLSRK 507
Db 261 MKYRGYDKRKLPLKLELEFASIMKELQLYEAEFT-----GYEIVDKHKTFELEIK 313
Qy 508 VKE-----KOTARD---MIVWIEETIYPVHGFEVTL-----TIVVQTLDOI 545
Db 314 LKEVPSFALALETSSLDPFNCEIVGISVSFKPKTAYYIPLHHRNAQNLDLTVLSKLKEI 373
Qy 546 GSKSTHVLTVLERYGQVFSKLPDNDKQVMLLSQVSTYWKNNVQMTAVAD----- 597
Db 374 LEDPSSKIVGQNLKY-----DYKVLVKGISPVYPHFDTMIAAYLLEPNKKFN 422
Qy 598 -----RMGYRLVSNQAIWRVW-----ESPENVDOFHVSDQWPEIL 633
Db 423 LEDLSLKFGLYKNTSYQELMSPFLGFSFADVPVDKKAANYSCEDAD---ITYRYLKIL 479
Qy 634 GNALNK-----TYNRISD-----LRKDISNITKNVLVAEKASANARVELEAAESKLSLV 682
Db 480 SMKLHEALENVFYRIEMPLVNLVARMELNGVYVDFELKLLSEEGYKGLLEAEKIYQI 539
Qy 683 EGEVPLGENPAKMKRL-----KSTVKTGTGAELSLRESLEAKEALLNRLALSETEV 733
Db 540 AGEPEINSPKQVSKILFEKLIKPRGKTKTG--EYSTR--IEVLEEIAN-----EHEIV 591
Qy 734 LLLLFQSLGLVKE-----RLPDPTKVRSVQDLKLSIGAEDDKPSAMD-----VDSE 779
Db 592 PLILEYRKIQKLSYIIDLTPKLVNPKTGRIHASFQGTGTATGRUSSDPNQLNPTKSE 651
Qy 780 NGNPKKSCVEGEREQWCLSTGLYLTAFTRQYASEIWPHM---EKLESEVFSGEDVHPL 834
Db 652 EGKEIRKAIVQDPDPFWIVSADY-----SQIELRLIHLSGDENLVKAFEEGIDVHTL 704

RESULT 8
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DAUBERSTES, PIERRE
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide

US-08-973-462-8
Query Match 2.7%; Score 116.5; DB 4; Length 1786;
Best Local Similarity 18.2%; Pred. No. 0.13;
Matches 84; Conservative 83; Mismatches 152; Indels 143; Gaps 18;
Qy 446 VQEILOREVLRSYDKIKOSI-EN-----ATALELLPPKAGPNFMYSLEEG 491
Db 699 VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPSVVESVAPSVEES 758
Qy 492 KEKTEEQOLSAELSRKVE-----KOTARDMIVMIEETIYPVHGFEVTLIVVOTLTD 544
Db 759 VEENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPSVVESVAPSVEES 818
Qy 545 TGSKSFTHLVTLVRYGVFSLKCPD---NDKQVMLLSQVSTYWKNNQMTAVADRMGM 601
Db 819 VATNLSDSLNLN--LGGTETEIKDSILNEIEEKENVVTITILENVEETTAESVTFN 876
Qy 602 Y-----RLVSNQATVRRVSPENVDQPHVSDQPHWELGNALNKTYNR-----ISD 646
Db 877 ILBEEQENTITNDTI-----BEKLEELHEN-----VLSAALENTOSEEKEVIDVIEE 925
Qy 647 LRKDI-SNITKNVLVAEKASANARVEL-----EAAESKLSLVEGSPVLGEN--PAKMKR 697
Db 926 VKEEVATTLTIEVQAEKSANTITEIFENLEENAVESNAENLEKLNIEVFNVTLDK 985
Qy 698 LKSTVEKTGEA-----EAKALNRA--LSETE-----VLLLLLLQSFGLGVK 714
Db 986 VEETVEISGESLENEMDKAFFSEIFDNVKGIOENLLTGMFRSIFTSIVIOSEEKVDLNE 1045
Qy 715 SL-----EAKALNRA--LSETE-----VLLLLLLQSFGLGVK 746
Db 1046 NVSSILDNIENKNGELNKLKNISSTEGVQETVTEHVQNVYVDVDPANKDQFLGIIN 1105
Qy 747 E-----RLPDPKTVRSVDOLKSIGAEDDKPSAMDYSE 779
Db 1106 EAGGLKEMFFNLEDVFK--SESDVITVEIKDEPVQKEVEKE 1145

RESULT 9
US-09-019-160-5
; Sequence 5, Application US/09019160
; Patent No. 6306588
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; APPLICANT: Solus, Joseph
; APPLICANT: Yang, Shuwei

; TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic
; TITLE OF INVENTION: Nucleic Acid Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,160
; FILING DATE: 06-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: (To be assigned)
; FILING DATE: 06-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/037,393
; FILING DATE: 07-FEB-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.4250002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-09-019-160-5

Query Match 2.7%; Score 116; DB 4; Length 708;
Best Local Similarity 18.4%; Pred. No. 0.029;
Matches 98; Conservative 93; Mismatches 188; Indels 154; Gaps 22;

Qy 422 QETWPEEWAFVLDLPKAPKRVFQVEILQREVRLSY-----WDKIK--- 463
Db 21 QLLGKYRNLEYILEHARELPQVRKALLRRDREVAIILSKKLATLVTNAPVEVDMEKRYG 80
Qy 464 -----QSTENATALEEL-LPPKAGPNFMYSLEEGKEKTEEQOLSAELSRKVE-- 510
Db 81 YDKRKLLPILKELEFASIMKELQLEYEAAPT-----GYEIVKDHKTFEDLIEKLEVP 133
Qy 511 -----KQIARD-----MIVWIEETIYPVHGFEVTL-----TIVVOTLTDIGSKSF 550
Db 134 SFALAELETSLDPNCEIGVSGISVSPFKPTAYIYIPLHNRNAQNLDLTVLSKLKEILEDPS 193
Qy 551 THLVTVLERYGVFSLKCPDNDQVMLLSQVSTYWKNNQMTAVAD----- 597
Db 194 SKIVGNLKY-----DYKVLVKGISPVYFPHFDTMIAAYLLEPNKFNEDLS 242
Qy 598 -RMMGYRLVSNQAIVRWV-----FSPENVDOFHVSDQPHWELGNALN 638
Db 243 LKFLGYKMTSYQELMSFSSPLFGFSADVPVDKAANYSCEDAD---ITYRLYKLSMKLH 299
Qy 639 K-----TYNRISD-----LRKDISNITKNVLVAEKASANARVELEAESKLSLVEGPV 687
Db 300 EAELENVYFRIEMPLVNLARMELNGVYVDTEFLKILSEEYGGKLEAEKIYQIAGEPF 359
Qy 688 LGENPAKMKRL-----KSTVEKTGEAELSLRESLEAKALLNRLSETETEVLLLLLF 738

Db 360 NINSPKOVSKILFEKLGKIPRGKTKTG--EYSTR--IEVLEEIAN-----EHEIVPLILE 411
QY 739 QSFLGLVKE-----RLPDPKTVRSVODLKSIGAEDDKPSAMD-----VDSSENGNPK 784
Db 412 YRKIOKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRSLSSDPNQLNLPKSEEGKEI 471
QY 785 KSCVEGREOWCLSTGLYLTAFTRQYASEIWPHM---EKLSEVFSGEDVHPL 834
Db 472 RKAIVPQDPDWIVSADY-----SQIELRILAHLSGDNLVKAFEEGIDVHTL 519

RESULT 10

US-09-019-160-6
; Sequence 6, Application US/09019160
; Patent No. 6306588
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; APPLICANT: Solus, Joseph
; APPLICANT: Yang, Shuwei
; TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic
; TITLE OF INVENTION: Nucleic Acid Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,160
; FILING DATE: 06-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: (To be assigned)
; FILING DATE: 06-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/037,393
; FILING DATE: 07-FEB-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.4250002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 893 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-09-019-160-6

Query Match 2.7%; Score 116; DB 4; Length 893;
Best Local Similarity 18.4%; Pred. No. 0.043;
Matches 98; Conservative 93; Mismatches 188; Indels 154; Gaps 22;
QY 422 QFTWPHWEAFVLDLPKWAQKRVFVQEILOREVRLSY-----WDKIK--- 463
Db 206 QLLGKRYNLEYILEHARELPQVRKALLRDREVAIILSKKLATLVNAPVEVDWEEMKRYG 265
QY 464 -----QSIENATALEEL-LPPKAGPNFMYSLSEEGKEKTEEQLSAELSRKYKE-- 510

Db 266 YDKRKLPLILKELEFASINKELQLYEEAPT-----GYEIVKDHKHTFEDLIEKLKEVP 318
QY 511 -----KOTARD---MTVWIEETIYPVHGFEVTL-----TIVVOTLLDIGSKSF 550
Db 319 SFALALETSSLDPFNCEIVGISVSFKPTAYVYIPLHRRNAQNLDLTVLSKLKEILEDPS 378
QY 551 THLVTVLERYGOVFSKLCPDNDKQVMLLSQVSTYHKN-NVQMTAVAI----- 597
Db 379 SKIVQONLKY-----DYKVLVYKGISPVYPHFDMTIAAYLLEPNEKKFNLEDLS 427
QY 598 -RMGYRLVSNQAIVRWV-----FSPENVDOFHVSQDPWEILGNALN 638
Db 428 LKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAAANYCEDAD---ITYRLVKILSMKLH 484
QY 639 K-----TYNRISD-----LRKOISNITKNVLVAEKASANARVELEAEKSLSLVEGEVP 687
Db 485 EAELENVFIYRTIEMPLVNLVARMELNGVYVDTFELKKLSEYGGKLEELAEKTYQIAGEPF 544
QY 688 LGENPAKMKRL-----KSTVEKTGEAELSRESLEAEKALLNRLALSETEVLLLLLF 738
Db 545 NINSPKOVSKILFEKLGKIPRGKTKTG--EYSTR--IEVLEEIAN-----EHEIVPLILE 596
QY 739 QSFLGLVKE-----RLPDPKTVRSVODLKSIGAEDDKPSAMD-----VDSSENGNPK 784
Db 597 YRKIOKLKSTYIDTLPKLVNPKTGRIHASFHQTGTATGRSLSSDPNQLNLPKSEEGKEI 656
QY 785 KSCVEGREOWCLSTGLYLTAFTRQYASEIWPHM---EKLSEVFSGEDVHPL 834
Db 657 RKAIVPQDPDWIVSADY-----SQIELRILAHLSGDNLVKAFEEGIDVHTL 704

RESULT 11

US-09-019-160-7
; Sequence 7, Application US/09019160
; Patent No. 6306588
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; APPLICANT: Solus, Joseph
; APPLICANT: Yang, Shuwei
; TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic
; TITLE OF INVENTION: Nucleic Acid Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,160
; FILING DATE: 06-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: (To be assigned)
; FILING DATE: 06-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/037,393
; FILING DATE: 07-FEB-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.4250002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 893 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-09-019-160-7

Query Match 2.7%; Score 116; DB 4; Length 893;
Best Local Similarity 18.4%; Pred. No. 0.043;
Matches 98; Conservative 93; Mismatches 188; Indels 154; Gaps 22;

```
QY 422 QFIWPEEWAFLDLPKAPKRVFQEILOREVRLSY-----WDKIK--- 463
DB 206 QLLGKYNLEYLHARELPQRVKALLRDREVAISKKLATLVNAPVEVDWEEMKYRG 265
QY 464 -----QSIENATALEEL-LPPKAGNFMYSLSEEGKEKTEEQOLSAELSRKVK-- 510
DB 266 YDKRLLPILKELEFASIMKELQYEEAPT-----GYEIVKDHKTFFDLIEKLEVP 318
QY 511 -----KOTARD-----MIVMIETIYPVHGFEVTL-----TIVVOTLLDIGSKSF 550
DB 319 SFALAEETSSLDPFNCEIVGISVSPKPTAYIPLHHRNAQNDELTVLSKLEILEDPS 378
QY 551 THLVTVLERYGOVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAI----- 597
DB 379 SKIVGNLKY-----DYKVLWVGISPVYPHFDTMIAAYLLEPNKKNFNLEDLS 427
QY 598 -RMGGRVLSNQAIVRWV-----FSPENVDFHVSQDPWEILGNALN 638
DB 428 LKFLGYKMTSYQELMSFSPFGFADVPVDKAANYSCEDAD---ITYRLYKILSMKHL 484
QY 639 K-----TYNRISD-----LRKDISNITKNVLVAEKASANARVELEAAESKLSVGEVP 687
DB 485 EAELENVYRIEMPLVNLARMELNGVYVDTEFLAKLSEEGYKKLEAEKIYQIAGEPF 544
QY 688 LGENPAKMKRL-----KSTVEKTGEAELSLRESLEAKALLNRLALSETEVLLLLLF 738
DB 545 NINSPKQVSKILFEKLGKIPRGKTKTG--EYSTR--IEVLEIAN-----EHEIVPLILE 596
QY 739 QSFGLVLKE-----RLPDPKTVRSVQDLKSGIADDDKPSAMD-----VDSSENGPK 784
DB 597 YRKIQKLKSTYIDTLPKLVNPKTGRHASFHTGTATGRUSSDPNQLNPTKSEEGKEI 656
QY 785 KSCVEGERQWCLSTGLYLTATQRYASEIWPHEM-----EKLESEVFGEDVHPL 834
DB 657 RRAIVPQDPDMWIVSADY-----SQTIELRLAHLSDGNLVKAFEGIDVHTL 704
```

RESULT 12
US-09-019-160-8
; Sequence 8, Application US/09019160
; Patent No. 6306588
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; APPLICANT: Solus, Joseph
; APPLICANT: Yang, Shuwei
; TITLE OF INVENTION: Polymerases for Analyzing or Typing Polymorphic
; TITLE OF INVENTION: Nucleic Acid Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/019,160
FILING DATE: 06-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA: (To be assigned)
FILING DATE: 06-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/037,393
FILING DATE: 07-FEB-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942.4250002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 893 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-019-160-8

Query Match 2.7%; Score 116; DB 4; Length 893;
Best Local Similarity 18.4%; Pred. No. 0.043;
Matches 98; Conservative 93; Mismatches 188; Indels 154; Gaps 22;

```
QY 422 QFIWPEEWAFLDLPKAPKRVFQEILOREVRLSY-----WDKIK--- 463
DB 206 QLLGKYNLEYLHARELPQRVKALLRDREVAISKKLATLVNAPVEVDWEEMKYRG 265
QY 464 -----OSTENATALEEL-LPPKAGNFMYSLSEEGKEKTEEQOLSAELSRKVK-- 510
DB 266 YDKRLLPILKELEFASIMKELQYEEAPT-----GYEIVKDHKTFFDLIEKLEVP 318
QY 511 -----KOTARD-----MIVMIETIYPVHGFEVTL-----TIVVOTLLDIGSKSF 550
DB 319 SFALAEETSSLDPFNCEIVGISVSPKPTAYIPLHHRNAQNDELTVLSKLEILEDPS 378
QY 551 THLVTVLERYGOVFSKLCPDNDKQVMLLSQVSTYWKNNVQMTAVAI----- 597
DB 379 SKIVGNLKY-----DYKVLWVGISPVYPHFDTMIAAYLLEPNKKNFNLEDLS 427
QY 598 -RMGGRVLSNQAIVRWV-----FSPENVDFHVSQDPWEILGNALN 638
DB 428 LKFLGYKMTSYQELMSFSPFGFADVPVDKAANYSCEDAD---ITYRLYKILSMKHL 484
QY 639 K-----TYNRISD-----LRKDISNITKNVLVAEKASANARVELEAAESKLSVGEVP 687
DB 485 EAELENVYRIEMPLVNLARMELNGVYVDTEFLAKLSEEGYKKLEAEKIYQIAGEPF 544
QY 688 LGENPAKMKRL-----KSTVEKTGEAELSLRESLEAKALLNRLALSETEVLLLLLF 738
DB 545 NINSPKQVSKILFEKLGKIPRGKTKTG--EYSTR--IEVLEIAN-----EHEIVPLILE 596
QY 739 QSFGLVLKE-----RLPDPKTVRSVQDLKSGIADDDKPSAMD-----VDSSENGPK 784
DB 597 YRKIQKLKSTYIDTLPKLVNPKTGRHASFHTGTATGRUSSDPNQLNPTKSEEGKEI 656
QY 785 KSCVEGERQWCLSTGLYLTATQRYASEIWPHEM-----EKLESEVFGEDVHPL 834
DB 657 RRAIVPQDPDMWIVSADY-----SQTIELRLAHLSDGNLVKAFEGIDVHTL 704
```

RESULT 13

US-08-353-700-1
: Sequence 1, Application US/08353700
: Patent No. 5599919
: GENERAL INFORMATION:
: APPLICANT: YEN, TIMOTHY J.
: APPLICANT: RATTNER, JEROME B.
: TITLE OF INVENTION: NUCLEIC ACID ENCODING A
: TITLE OF INVENTION: TRANSLIENTLY-EXPRESSED KINETOCHORE PROTEIN,
: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
: STREET: 1601 MARKET STREET, SUITE 720
: CITY: PHILADELPHIA
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/353,700
: FILING DATE: 09-DEC-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: REED, JANET E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: HUMAN
US-08-353-700-1

Query Match 2.6%; Score 113.5; DB 1; Length 3248;
Best Local Similarity 18.7%; Pred. No. 0.77;
Matches 187; Conservative 139; Mismatches 332; Indels 341; Gaps 45;
QY 9 LRIGKPEVGTSSD-YKDIETCFGVRIRIERSQGVLPFLQC-----A 54
DB 2232 LHAIEKLERERENDSLKDYEN---LERELQSEENQELVILDAENSKAEVETLKTQI 2287
QY 55 EOLPHKIPLYGTGLGLNLENEFVQKLVSFVHANFOALDNGNCSIRILLRFTWLSLC 114
DB 2288 EEMARSKIFELDVTLRSEKNTKQIEK-----QGLSELKLLSLFKSLLE 2337
QY 115 SKVIQASLIVFETLLSSAATTVEEKGPNQWQADFYVICILSSLPWAGSELAE--- 171
DB 2338 EK-EQA-----EIQIKESKTAVEMLQNLKELNEVAALC-----GDQIMKATE 2382
QY 172 -----QVPEIERVLVIGQAYLSIRKNSSTGLNFFHNGEFESSLAERDFVEDL 220
DB 2383 QSLDPPIEEHQLRNLSIEK---LRARLEADEKKQLCVLQQLKESEHADLL-KGRVENL 2437
QY 221 LDRIQSLASNGWKLSEVPRHLSFEAQLVACKFHELR-PTKCMQPPSPDHSRAYSQK 279
DB 2438 ERELEIARTN-----QEHAALAEANSKGEVETLAKIBGMTQSLRGLDVLVVTIRSE 2489
QY 280 KH---DALTRYPQIRRL-----NIFPANKMEDVQIDRFVVEEYLLDVLFLNGCR 328
DB 2490 KENLTNELQKEQERISELEINSSFNILQKEQKVQ-----MKEKSSTAMEMLOTQL 2543

RESULT 14
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Rattner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700

FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 2.6%; Score 113.5; DB 5; Length 3248;
Best Local Similarity 18.7%; Pred. No. 0.77;
Matches 187; Conservative 139; Mismatches 332; Indels 341; Gaps 45;

Qy 9 LRIGKGPYGTSSD-YKHITCTCGVIRRETERSGDQVLPFLQC-----A 54
Db 2232 LHIAEKLKERENDSLDKVEN-----LERELQMSSENOELVILDAENSKAEVETLKTQI 2287
Qy 55 EOLPHKIPLYGLIGLNLNENEDFVKLVESVHANFOVALDSGNCNSIRILLRFMTSLIC 114
Db 2288 EEMARSLKIFDLVTLRSEKENLTKQIEK-----OQJSELDKLLSSKSLLE 2337
Qy 115 SKVIOPASLIIVFETLLSSAATVDEEKGNSWPQADPVYICILSSLPWGGSELAEE--- 171
Db 2338 EK-EQA-----ETQIKESKTAVEMLQNLQKELNEVAALC-----GQEIINKATE 2382
Qy 172 -----QVPEIERVLVIGIAYLSIRKNSSTSGLNFNHNGEFESSLAEKDFVEDL 220
Db 2383 OSLDPPIEEHQRLRSIEK-----LRLAEADKKOLCVLQQLKESEHHADLL-KGRVENL 2437
Qy 221 LDRIQSLASNGKLESVPRPLSFEALVAGKFHEL-RPIKMEQPPSPSDHSRAYSQK 279
Db 2438 ERELEIATN-----QEHAALEAENSKEGVEYTLKAKIEGWTQSLRGLDLDVVTIRSE 2489
Qy 280 KH---DALTRYPQIRRL-----NIFPANKMEDVQPIDRFVVEYDLYLVFLNGCR 328
Db 2490 KENLTNELOKEQERISELEIINSFPENILQEKQVKQ-----MKESSSTAMEMLTQL 2543
Qy 329 KECASYMANLPVTFRYEYLMATLFSQILLPQPFKTLTYLVIMDLCKALPGAPPA-- 386
Db 2544 KENLNRVAALHNDQACKAKQONLSQVECLE-----LEKAQLLQGLDEAKN 2590
Qy 387 ---VVAGAVRALFEKISD-----LDMESRRLI-----LWFSH 416
Db 2591 NYIVLOSSVKGLIQVEGKQKLEKDEISLKNQIQDQOBOLVSKLSQVEGHEQLWKQ 2650
Qy 417 HLSNPFQIWPMEWAFVLDLPKAPKRVFVQEI-----LOREVLSYWDKIK--Q 464
Db 2651 NLELRLNLTVEEQIKQVLOQ-----SKNASLODTLEVLOSSYKNLENELELTRKMDMSFEV 2705
Qy 465 SIENATALEELLPPKAGPNFYSLSEGEKKT-----EQQISAELSRKVKKEKQATAR 515
Db 2706 KVNKMTAKETEL-----QREHMAQKTAEOBELSGEKNRLAGELQLLEETKSSK 2757
Qy 516 DMIVWIEETIPVHGFVTL-TIVVOTLLDIGKSFTHLVTVLRYGVQVFSKLC----- 568
Db 2758 DQLK-----ELTLENSEKSLSDCMHKD-----QVEEGKVRREEIAEYQLRL 2799
Qy 569 ---PDNDKQVMLLS-----QVSTYKNNVQMTAVAI DRMGYRLVSNQAIVRWVFSPE 619
Db 2800 HEAEKKHQALLDTNKOYEVEIQTYRE-----KLTSEKEC----- 2834
Qy 620 VDOFHVSQDPWEI-----LGNALNKTYNRISDLRKDISNITK--NLVVAEKASANA 668

Db 2835 -----LSSQKLEIDLLKSSKEELNLSLKATTOILEELKTKMDNKLKYVNQLKKENERAQG 2889
Qy 669 RVEL-----EAAESKLSLYEG-----EPVLGENPAKMKRLKSTV-EK 704
Db 2890 KMKLLKSKQLEEREIIOKELSQLOAAQEKOKGTGVMDTKVDELTEIKELKETLEEK 2949
Qy 705 TGEAE-----LSLRESLE-AKALLNRLALSETEVLLLLLFQS-----FLGVLK 746
Db 2950 TREADEYLDKYCSLLISHEKLEKAKEML-----ETQVAHLCSQOQSKQDSRGSPLPGPV- 3002
Qy 747 ERLPDPKTVRSVODLKSIGAEDDKPSAMDVDSNGNPNPKSCVEGREQWLSTLGLYTAF 806
Db 3003 --VPGSPIPSVTE-KRLSSGQNKASG-----MEKLESEVFSGEDVHP 833
Qy 807 TRQYASEIWPH-----MEKLESEVFSGEDVHP 833
Db 3027 KRQRSSGIWENGRTPTPATPESFSKSKKAVMSG--IHP 3063

RESULT 15

US-09-356-952-2

Sequence 2, Application US/09356952

Patent No. 6117663

GENERAL INFORMATION:

APPLICANT: Boriack-Sjodin, Ann

APPLICANT: Margarit, S. M.

APPLICANT: Bor-Sogli, Dafna

APPLICANT: Cole, Philip

APPLICANT: Kuriyan, John

TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

FILE OF INVENTION: THEREOF

FILE REFERENCE: 600-1-228N

CURRENT APPLICATION NUMBER: US/09/356,952

CURRENT FILING DATE: 1999-07-19

EARLIER APPLICATION NUMBER: 60/093,631

EARLIER FILING DATE: 1998-07-21

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 1333

TYPE: PRT

ORGANISM: Homo sapiens

US-09-356-952-2

Query Match 2.6%; Score 112; DB 3; Length 1333;
Best Local Similarity 19.5%; Pred. No. 0.22;
Matches 140; Conservative 97; Mismatches 200; Indels 280; Gaps 38;

Qy 273 RAYSGKQKHDAITRYPQIRRLNIFPAN-----KMEDV-QPIDRFVVEYLLDV----- 320
Db 406 REYSQOMKQOLA-----IKKXNEIOKNDGWEGKDQGCCNEFIMEGTLTRVGAKHERH 460
Qy 321 LFLYNGCRKECASY--MANLPVTFRYEYLMATLFSQILLPQPFKTLTY--YTLVIMD 375
Db 461 IFLFDGLMICCKSNHGOPLPGASNAEYRLKKEFFMRKVQINDKDDTNEYKHAFEILKD 520
Qy 376 LCKALPGA-----FPAVAVAGAVRALFEKISDLDMSRTRLILWFSHLSNFQFIWP 426
Db 521 ENSVIFSAKSAEKNWMAALISLQYRSTLERMLDVTM----- 558
Qy 427 WBEWAFVLDLPKAPKRVFVQEILOREVLSYWDKIK-----QSTENATALEELLPPKAG- 481
Db 559 -----LOBEKEQOMRLPSADVIRFAEPDSEEN-IIEENMQPKAGI 598
Qy 482 -----PNFM-----YSLEEGKEKTE 496
Db 599 PIKAGTVIKLIERLTFTYHYADPNFVTFLTITRYSFCKPOELLSLITERFEIPE-PEPTE 657
Qy 497 -----EQQLSAELSRKVKKEKQATDMIV-----WIEETIYPVHGFVTLTIVVQT 541
Db 658 ADRIATENGDPULSAELKRFKEYIQVQLRVNLVNCRHWHYEHFY---DFE----- 705

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:51:51 ; Search time 63 71 Seconds
(without alignments)
1478.427 Million cell updates/sec

Title: US-09-882-986-2
Perfect score: 4374
Sequence: 1 MSNKTLLLRIGKGPVGT.....EDVHPLFLQAISSALQPLH 848

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	818.5	18.7	800	22	ABB62896 Drosophila melanog
2	550.5	12.6	747	22	ABB63581 Drosophila melanog
3	145.5	3.3	1833	22	ABB71141 Drosophila melanog
4	141.5	3.2	5373	22	AAU14603 Novel bone marrow
5	140	3.2	906	22	AAG67251 Amino acid sequenc
6	140	3.2	906	22	AAU14603 Human protein sequ
7	138.5	3.2	3868	22	ABB63663 Drosophila melanog
8	137.5	3.1	814	22	AAU93442 Human polypeptide,
9	135	3.1	3854	22	ABB67104 Drosophila melanog
10	135	3.1	5385	22	ABB66487 Drosophila melanog
11	135	3.1	5496	22	ABB67161 Drosophila melanog

12	135	3.1	8805	22	ABB67112 Drosophila melanog
13	134.5	3.1	1084	22	AAU93096 Human protein sequ
14	133.5	3.1	710	22	AAU78497 Human protein SEQ
15	133.5	3.1	3263	22	ABB67210 Drosophila melanog
16	132	3.0	1743	19	AAU98879 H. pylori GPO 175
17	131.5	3.0	2154	20	AAU81639 Mouse elf-1 protei
18	131	3.0	2633	22	ABG05505 Novel human diagno
19	131	3.0	2688	22	AAU40883 Human polypeptide
20	131	3.0	5447	22	AAU14697 Novel bone marrow
21	130	3.0	6815	22	ABB66811 Drosophila melanog
22	129.5	3.0	981	22	AAU36618 Staphylococcus aur
23	129.5	3.0	2272	18	AAU21731 GAL4/HA/NuMA fusio
24	129	2.9	2663	22	AAU33097 Human polypeptide
25	128	2.9	4590	22	AAU33184 Novel human secret
26	128	2.9	7201	22	ABU71136 Drosophila melanog
27	127	2.9	980	22	AAU62027 Recombinant P. fur
28	127	2.9	1003	20	AAU34599 Chlamydia pneumoni
29	127	2.9	4536	19	AAU41262 Apolipoprotein B-1
30	126.5	2.9	808	22	ABG05140 Novel human diagno
31	126.5	2.9	1269	22	AAU40297 Human polypeptide
32	126.5	2.9	2013	22	ABB62322 Drosophila melanog
33	125.5	2.9	2415	22	ABG20279 Novel human diagno
34	125	2.9	469	19	AAU37881 BRCAL modulator pr
35	125	2.9	469	20	AAU30149 Amino acid sequenc
36	125	2.9	1671	11	AAU06341 Signal peptide and
37	125	2.9	1829	18	AAU29322 DNA polymerase wit
38	124.5	2.8	2400	22	ABG20278 Novel human diagno
39	123	2.8	677	22	AAU85608 Secreted protein c
40	123	2.8	677	22	AAU85608 Human myosin heavy
41	123	2.8	809	20	AAU09371 Human apolipoprote
42	123	2.8	1303	22	ABB62088 Drosophila melanog
43	123	2.8	2721	8	AAU70647 Sequence of N-term
44	123	2.8	4536	20	AAU96826 Amino acid sequenc
45	122.5	2.8	1521	21	AAG39235 Arabidopsis thalia

ALIGNMENTS

RESULT 1
ABB62896
ID ABB62896 standard; Protein; 800 AA.
XX
AC ABB62896;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 15480.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
(PEKE) PE CORP NY.
Venter JC, Adams M, Li PWD, Myers EW;
WPI; 2001-656860/75.
N-PSDB; ABL06999.
New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

Qy	115	SKVTPASLIVVFETLSSAATTVD--EEKGNPWSQPADFYVICILSSLPFWGSELAEOV	173
Db	137	cqvipssyikllaafikdcealkdpddlvgtvpqrrdflaycvisampllgdrldgeet	196
Qy	174	PDETERVLVQIAYLSRTRKNSSTSGLFFNFGFESSLAEKDVEDLLDRIOSLASNGW-	232
Db	197	--afdklvislqlyikrsalhtnmsvw-----pdfnqrldylellwqvgdgmqrhwa	248
Qy	233	--KLESVPRLHSPEAQLVAGKTFHELRIPIKCMQEPSPPSDHSRAYSQKQKHDAALTRY	290
Db	249	epehqliprykfsfsetlsgrihqlr-----dydla-----aheercryplp	291
Qy	291	IRRLNIPANKMEDVQ-----PIDRVEVEYLLDVLFLNGCRKECAS-----YMANLPV	340
Db	292	rvcfriifscdsvgeipnmpppvsvlerhlleahildilisfhkerkicadsllymyaaskp	351
Qy	341	TFREYELMAETLFSQILLPQPFPKTLTYTLVTIMDLCKALPGAFPAVAVAGAVRALFEKIS	400
Db	352	lpvy-ycivevilgemrlpttanwstlaygsilveickrqdpkipqvavaqaidlynrln	410
Qy	401	DLDMESTRLILMFSHLSNFQFTWEEWAFVLDLP-----KWAPKRVFVOETILOREV	454
Db	411	smsvacfdrlvnnvshisnfgncqskwaqglpsdpdsatnlqpkvvfirellkkcf	470
Qy	455	RLSWDKIKOSIENATALEELLPPKAGPNPMWSLEBEGEKTEBQQLSAELSRKVKEKOTA	514
Db	471	r-----ikdvvpd--lladflppvpvlphfkfv-----delpgailskdlleam:spqas	518
Qy	515	RDIMVTEETIYPVHGFEVTLTIIVQTLLDIGSKSFTHLVTLVRLYGVQVFSKL-CPDN	573
Db	519	pemselikstgti-gpllxinvtqncnlhgsksfshfgilrkyhsvfkdlvagdp	577
Qy	574	QVMLLSQVSYTWKNNVQTAVALDRMGVRLVSNQAIVRWVFSPP-----ENVDOFHVSDQ	630
Db	578	qaavlngifdvwvasdykfvtckivtgliliepvisvrvifgspmrkeltkiyi-----w	633
Qy	631	EILGNAL 637	
Db	634	ellhsal 640	
RESULT 3			
ABB71141			
ID	ABB71141	standard; Protein; 1833 AA.	
AC	ABB71141;		
XX			
DT			
XX			
DE	26-MAR-2002	(first entry)	
XX			
XX	Drosophila melanogaster	polypeptide SEQ ID NO 40215.	
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.		
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US09231.		
XX			
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX			
XX	(PEKE) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
XX	WPI; 2001-656860/75.		
DR	DR		
XX	N-PSDB; ABL15244.		
XX			


```

969 l rqlvmeta pvs gl -- les i qkanell ediat gvs nylekkrly f p-----rf 1015
Qy 715 SLEAK EALLN RALST E V L L L L-----FQSLGVLKERLPDP T K VRS VQD 760
Db 1016 fflandem l -e ll set k d p l r v p h l s k c f e g i n s l e r d a a k n v l a m i s s k t e f i e q 1074
Qy 761 LKS-----IGADGDPSKMDVDS E N C---NPKKSC E V G E R Q W C L S T L G Y L T A F 806
Db 1075 v s t a a a g g s e k w l i g v e d e m l k a v r y q n e l s f a h y p k v-----k r h e w v l e-----1121

Qy 807 T R Q Y A S E I W P H M E K L E - S E V F S G E D V H 832
Db 1122 -----w o o m t u l a t e c v v u a s r v b 1140

```

```

RESULT      8
AAM93442
ID AAM93442 standard; Protein; 814 AA.
XX
AC AAM93442;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide, SEQ ID NO: 3084.
XX
XX Human: full length cDNA; cDNA synthesis; oligo-capping.
XX
XX Homo sapiens.
XX
XX EPI130094-A2.
XX
XX 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX
XX 11-JAN-2000; 2000JP-0118774.
XX
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI: 2001-524255/58.
XX
XX N-PSDB: AAK94363.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 3084; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 814 AA;

Query Match 3.1% Score 137.5; DB 22; Length 814;
Best Local Similarity 19.9%; Pred. No. 0.021;
Matches 108; Conservative 92; Mismatches 244; Indels 99; Gaps 24

QY 341 TFRYEYLMAEFLFSQILLPOP--PFKTLTYTL--VIMDLCKALPG-AFPVVVAG----- 390

```

```

Db 277 sirysgsskvtlmalclmalpfipasnlfifvgfvaervlypsmgfcilvahgwqk 336
Qy 391 AVRALFEKISOLDMESRTRLLWFSHHLSNFQFIWPEWAFVLDLPKAPKRVFVQFIL 450
Db 337 stksvfkkskicis-----nvilthsktthrnwdw-----seytlfmsalk 380
Qy 451 QREVLRSVWDKIKOSIENATALEEL-----PKAGPNF-----MYSLEEGREKT 495
Db 381 vknknaklwnvghalenekneferalkyflgathvqddigahnmvgrtyknlnrtkeae 440
Qy 496 EEOQLSAELSRKV--KEKOTARDMIVWTEETIYPVHGPEVTLTIWQFLDIDGSKSFTHL 553
Db 441 esymmakslmpqilpgkkyar-----iapnh-----lnvyi-nlanlianesrl 485
Qy 554 VTVLERYGVFSKLCPCDNDKQV-----MLLSQVSTYWKNNVQMTAVAIIDRMGMGYRLVSN 607
Db 486 eeadqlyrqais-mrpd-fkdayisrgellikmnkpikakeaylkaleldr-----n 535
Qy 608 QAIVRWVFPSPENVQDFHVS-DQPWEILGNALNKTYNRISDLRKDISNITKNVLVAEKASA 666
Db 536 nadi-ly-----nlaivhielkepeal-----knfnralelnphkhalfnlsaimvgesg 585
Qy 667 NARVELEAAESKLSLVEGEFVLGENPAKMRLKSTVEKTGTAELSLSRESL-----EAKCAL 722
Db 586 evklrpearckrlisyeenpidangyfnlgnlamddkdnneaewmkkaiklqadfsal 645
Qy 723 LNRALSTEVLLLLLFSFGVLKRLPDPTK-----VRSVDLKSIGAEDDKPSAMD 775
Db 646 fnlallysqtakelkaipileellryyphdikgillkgdlmngkdkilgakkcferile 705
Qy 776 VDSNGNPKKS-CEVGEREQWCLSTLGYL--TAFTRQVASEIWPHEKLESEVSGEDVH 832
Db 706 mdpsnvvgghnvcvvyfeekdlkaerclletlalapeehyiqrhlnivrdkissfsie 765
Qy 833 PLF 835
Db 766 pif 768

RESULT 9
ABB67104
ID ABB67104 standard; Protein: 3854 AA.
AC ABB67104;
XX ABB67104;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 28104.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL11207.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell

```

```

PT interactions -
XX Disclosure; SEQ ID NO 28104; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3854 AA;

Query Match 3.1%; Score 135; DB 22; Length 3854;
Best Local Similarity 19.1%; Pred. No. 0.46;
Matches 183; Conservative 156; Mismatches 348; Indels 270; Gaps 43;

Qy 55 EQLPHKIPLYGTLTGLLNEDEYQKLVSFVHAFQVALDSGNCN--SIRILLRFMTSL 112
Db 2186 erllqisiaehkplldklnktgealgalvaddgakinelltdcnaryaalrielterqga 2245
Qy 113 LCSKVQIPASIVVFETLLSSAATTVEBEKNPSWQPADVFYICILSSLPWGGSELAEQ 172
Db 2246 lesalqessqfsdklegmlralantvdqvn-----qld-----plsalp---qkireq 2290
Qy 173 VPDE-----IERVL-----VGIOAYLSIRK-NSTSGLNFFHNGEPES 209
Db 2291 iedndalmdldlkrqdafsvagraandviakagnkadvordikakleklnlnwdvqna 2350
Qy 210 SLAEKDFVEDLLDRIQSLASNGWK--LESVPRPHLSFEALVAGKPHLRPIKCMEQP-SP 267
Db 2351 tkkrqssldil---svaefwqqlnsvmtkldleel-----scqepaaq 2395
Qy 268 PSDHSRAYSGKQ--KHDALTRYQRIRLNIPPAKMDVQPIDRFVVEEYLLDV----- 320
Db 2396 pqdikkkqvalqelrheidqtke-veqvrhgsnlmmmcgpepkpevkhhledidnawd 2454
Qy 321 -LFYLGCRKECASYMANLPVTFRYEYLMAETLSQILLPOPPFKTLTYTVIMDLCKA 379
Db 2455 nitalyakree-----nlidamekamefhetiqnllfltkaedkfahlgavgsdi--- 2505
Qy 380 LPGAPPVAVGAVRALFEKISD-----LDMESRTLLILWFSHHLSNFQFIWPEEW 430
Db 2506 -----davkrqieqlksfkdevqphmveealnraqvelterts-----peq 2547
Qy 431 AFVLDLP-----KW-APKRVFVQEIQLQREVRLSWDKIKSI-----ENATALEEL 475
Db 2548 aasireplsvvnrtrweallrgmverqkglehallhlgfghalnellwinktdstldql 2607
Qy 476 LPPKAGP-----NFMYSLEE-GKEKTEEOQLSAELS-----RK 507
Db 2608 kpipgdpdllelvelaklvlandiqahqnsvdclndagrlletekgsveasttqeklrk 2667
Qy 508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIWVQTLTDI-----GSKSFTHL----- 553
Db 2668 lnnewkqlkqasdrqhealealreahgyiaevqilgldgvdavigaskpvggipeta 2727
Qy 554 VTVLERYGVFSKLCPCDNDKQVMLLSQVSTYWKNNVQM-----T 592
Db 2728 teglerfmevneidenrpkvetidagqgqeyikrqnmkvssnliqhtlrltikqrwdavv 2787
Qy 593 AVAIDRMGMGYRLV-----SNQAIVRWVFPSPENVQDFHVSQDPWEITLGNALNKTYNR 643
Db 2788 srasdkkikleialeatefhdtdiqafvewltqae-----kilsna--epvsr 2833
Qy 644 ISD-----LRKKDISNITKNVLVAEKASANARVELEAAESKLSLVEGEFVLGENP 692

```


KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
XX WO200171042-A2.
PN 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
PR (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL11264.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 28275; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABB01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5496 AA;

Query Match 3.1%; Score 135; DB 22; Length 5496;
Best Local Similarity 19.1%; Pred. No. 0.82;
Matches 183; Conservative 156; Mismatches 348; Indels 270; Gaps 43;

QY 55 EQLPHKIPYGLTGLNLENEDVOKLVESVHANTQVALDSCN--STRILLRWTSL 112
DB 3828 erllqsaehkplldkintgtgealgalvaddgdgakinelltdnaryaalrlrqrqa 3887
QY 113 LCCKVIQASLIWVFETLSSAATTVEEKGNSWQPOADFYVICILSSLPWGSSELAEO 172
DB 3888 lesalqessqfsdklegmralantvdqvn-----qld-----plsaip---qkireq 3932
QY 173 VPDE-----IERVL-----VGIOAYLSIRK--NSSTSGLNFFHNGEFES 209
DB 3933 iedndalmddldkrqdafasvqraandviakgnkdpavrdikakleklnnlwvndvqna 3992
QY 210 SLAEKDFVEDLDRIQSLASNGWK--LESVPRPHLSFEAQVLVAGKFHELRIKMEQSP-SP 267
DB 3993 tkkrgsglddl-----svaepfukqinsvmktlkdeetl-----scqeppaaq 4037
QY 268 PSDHSRAYSQKQ--KIDALTRYQIRRLNIFPANKMEDVQIDRFVVEYLLDV----- 320
DB 4038 pddikqgvalgeirheidqtke-veqvrrhgsnlmmmcgepdkpevkkhiedldnaw 4096
QY 321 -LFYLNCRKECASWANIYPTERYEVLMAETLFSOILLLPQPPFKTLYYTLVIMDLCKA 379
DB 4097 nitalyakree-----nlidamekamefhetlqnllkfltkaedkfahlgavgsdi--- 4147
QY 380 LPGAFPAVAVAGAVRALFEKISD-----LDMESRTRLILWFSHHLSNFQIWPWEW 430
DB 4148 -----davrqrleqlksfkdevdphmvevealnraqavelerts-----peq 4189

QY 431 AFVLDLP-----KW-APKRVFVQEIQLQREVRLSYWDKIKQSI-----ENATALEEL 475
DB 4190 aasireplsvvnrweallrgmverqkqlehallhlgqfghallnellwinktdstldql 4249
QY 476 LPPKAGP-----NPMYSLEE-GKEKTEEQQLSAELS-----RK 507
DB 4250 kplpgdpqlllevelaklvlandiqahqnsdvtindagrqlietckgsveastqeklrk 4309
QY 508 VKEK-----QTARDMIVWIEETIYPVHGFEVTLTIIVQTLDDI-----GSKSFTHL 553
DB 4310 innwkgllqkadsrghlealeareahgyaevqdlilgwldvavagaskpvggipeta 4369
QY 554 VTVLRYGVFSLKCPDNDKQVMLLSQVSTYWKNNVQM-----T 592
DB 4370 teqlerfmevyneldenrpkvvetiqagqgeyikrqgmkvssnqlqhtlrltkgrwdavv 4429
QY 593 AVAIDRMGMGYRLV-----SNQAIVRWVFSPENVDQFHVSDOPWEILGNALNKTYNR 643
DB 4430 srasdkkikieialkeatefhtlqafewltqae-----kllsna--epvsr 4475
QY 644 ISD-----LRKDISNITKNVLVAEKASANARVELEAESKLSLVEGEPVLGNP 692
DB 4476 vletiqagmeehkvlgkdvsthreamlildkkgth----lkysqkqdvfilknllvsvq 4531
QY 693 AKMKRLKS-TVEKTGEAELSRESLEAKEAL--LNRALSETEVILLLLFQ----- 739
DB 4532 hrervvksaaetraldhgykearefndawgmngylqeteqvldqileeataskpqq 4591
QY 740 --SFLGVLKE-----RLPDPTKVRSVQDLKSIGAEDDKPSAMDVDSENGPKKSCE 788
DB 4592 ikkyigklkethrqlgakqsvydgk-mrtgknlllerapkgdrp-vld-----kmlie 4641
QY 789 VGREWQCLSTGLYLAFTROYASEIWPHEKLESEVFSGEDVHPLFLOAISSALQF 845
DB 4642 l--keqw-----trvaksidrqrkleaalllsqg-----fsdaigellwd 4680

RESULT 12
ABB67112
ID ABB67112 standard; Protein; 8805 AA.
XX
AC ABB67112;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 28128.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL11215.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

PS Disclosure: SEQ ID NO 28128; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 8805 AA;

Query Match . 3.1%; Score 135; DB 22; Length 8805;

Best Local Similarity 19.1%; Pred. No. 1.8;

Matches 183; Conservative 156; Mismatches 348; Indels 270; Gaps 43;

QY 55 EQLPHKIPYGLIGLLNENEDFVKLVESVHANFOVALDSNCN--STRILLREWTSL 112

DB 7132 erllqsaehkplldkntgealgalvaddgdakineiltdnaryaalrlleirergga 7191

QY 113 LCSKVIQPASLIWVFETLLSSAATTVDDEKGNPSWQPADFFVICILSSLPWGSSELAEQ 172

DB 7192 lesaigessqfkdlegmlarantvdqvn-----qid-----plsalp--qkireq 7236

QY 173 VPDE-----IERVL-----VGIOAYLSIRK-NSSTSGLNFFHNGEFES 209

DB 7237 iedndalmddldkrqdfasvqraandviakgnkdpavrdkakleklnlnwvndvqna 7296

QY 210 SLAEKDFVEDLDRIQSLASNGWK-LESVPRPHLSFAQLVAGKFHELRIKMEQP-SP 267

DB 7297 tkkrgsslddl-----svaepfwkqlnsvmktlkleetl-----scqepaaq 7341

QY 268 PSDHSRAYSGKQ-KDALTRIPQRTLRRLNIFPANKMEDVQPIDRFVVEYLLDV----- 320

DB 7342 pqdikkqvalqeirheidqtkpe-veqvrirghsnmmcgpepdkpavkhhiedlnawd 7400

QY 321 -LFYLNCRKECASYMANLPVTFRYEYVLAETLFSOILLPQPEKTYLYTLMIDLCCKA 379

DB 7401 nitaiyakree-----nldamekamefhetlqnllkftkaedfahlgavgsdi--- 7451

QY 380 LPGAPVAVGAVRALFEKISD-----LDMSRTRLILWFSSHLSNFQFIWPEEW 430

DB 7452 -----davkrqieqlksfkdevphmvevealnqgavelterts-----peq 7493

QY 431 AFVLIDLp-----KW-APKRVFQEILOREVRLSYWDKTKQST-----ENATALEEL 475

DB 7494 asisreplsvvnrreallrgmverdkqiehalhlgqfghalnellvwinktdstldql 7553

QY 476 LPPKAGP-----NFMYSLEE-GREKTEEOOLSAELS-----RK 507

DB 7554 kplpdpqlllelvelaklkvlndiqahqnsvdclndagrqllieteksgveastqeklrk 7613

QY 508 VKEK-----QTARDMTVIEETIYPVHGFEVTLTVVQVTLDDI-----GSKSFTHL----- 553

DB 7614 lnnewkqlqkasdrqhelealeareahgylaevqdlilgwldgdavigaskpvggipeta 7673

QY 554 VTVLERYGOVFSKCLPDNDKQVMLLSOVSTYKNNVQM-----T 592

DB 7674 teqlerfmevyneldnrpkvetiqagqgeyikrqmqmkvssnqlghtlrltkqrwdavv 7733

QY 593 AVAIDRMCMYRLV-----SNOAIVRWVFPSPNDQFHVSDQWEILGNALNKTYNR 643

DB 7734 srasdkkilelalkatefhdclqafvewltqae-----kllsna--epvsr 7779

QY 644 ISD-----LRKDISNITKNVLVAEKASANARVELEAAESKLSLVVEGEPVLGENP 692

DB 7780 vlietlqameehkvldkdvsthreamlldkkgth-----lkyfsqkqdvllknlvsvq 7835

QY 693 AKMKRLKS-TVEKTGEAELSRESLEAKEAL--LNRALSETEVLLLLLFQ----- 739

DB 7836 hrwervvskaaertraldhgykearefndawsgmmgylqeteqvdqdlieeataskpqk 7895

QY 740 --SFLGVLYKE-----RLPDPFKVRSVODLKSIGAEDDKPSAMVDSDNGNPKKSCE 788

DB 7896 ikkyigklkethrqlgaksqsvydgtr-mrtgknlllerapkgdrp-vld-----kmile 7945

QY 789 VGREQWCLSTGLYLTAFTRQYASEIWPHEKLESEVFSGEDVHPLFLOATSSALQOF 845

DB 7946 l--kegw-----trvwsksidrqrklealllsgq-----fsdaigellidw 7984

RESULT 13

AAB93096

ID AAB93096 standard; Protein; 1084 AA.

XX

AC AAB93096;

XX

AC AAB93096;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:11947.

XX

DT Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

XX

PR 27-AUG-1999; 99JP-0300253.

XX

PR 11-JAN-2000; 2000JP-0118776.

XX

PR 02-MAY-2000; 2000JP-0183767.

XX

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX

PT full-length cDNAs defined in the specification, and for the detection

XX

PT and/or diagnosis of the abnormality of the proteins encoded by the

XX

PT full-length cDNAs -

XX

PS Claim 8; SEQ ID 11947; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesising 5602

XX

CC full-length cDNAs defined in the specification. Where a primer set

XX

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

XX

CC to the complementary strand of a polynucleotide which comprises one of

XX

CC the 5602 nucleotide sequences defined in the specification, where the

XX

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX

CC of an oligonucleotide comprising a sequence complementary to the

XX

CC complementary strand of a polynucleotide which comprises a 5'-end

XX

CC sequence and an oligonucleotide comprising a sequence complementary to a

XX

CC polynucleotide which comprises a 3'-end sequence, where the

XX

CC oligonucleotide comprises at least 15 nucleotides and the combination of

XX

CC the 5'-end sequence/3'-end sequence is selected from those defined in

XX

CC the specification. The primer sets can be used in antisense therapy and

XX

CC in gene therapy. The primers are useful for synthesising polynucleotides,

XX

CC particularly full-length cDNAs. The primers are also useful for the

XX

CC detection and/or diagnosis of the abnormality of the proteins encoded by

XX

CC the full-length cDNAs. The primers allow obtaining of the full-length

XX

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

cc represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1084 AA;

Query Match	3.1%;	Score 134.5;	DB 22;	Length 1084;
Best Local Similarity	17.9%;	Pred. NO. 0.061;		
Matches 162;	Conservative 143;	Mismatches 307;	Indels 293;	Gaps 37;
Qy	34	VIRREIERSGOVLPFLQCAQOLPHKPLVYGTLLI-----GLLNLENEDF	78	
Dy	265	vlnrvtrasspishfl-----dlisnvmvaplqscskvteafdylsflpigt---	316	
Qy	79	VOKLVESHANFOVALDNGCNSIRILRFTSLSCSVIQIPASLIVVFETLLSSAATV	138	
Dy	317	vqrlkavqplkvsmrmdc-----liivrkamfangldarksavagflilk-----	366	
Qy	139	DEEGNPSQPOADPYVICIISLSPWGSSELAEOVPDEIERVLVGIQAYLSIRKNSSTG	198	
Dy	367	-----nfkvlgslss-----sqcsqsl--svsqvhdvshy-----	396	
Qy	199	LNFPHNGEFESLAEKDFVEDLLDRIQSLASNGWKLSEVPRLHSEFAQLVACKHELR-	257	
Dy	397	-----hsvaneticfclmdslrclsq-----qadvrtilyegfydvirr	436	
Qy	258	-----PIKMEOPSPSDHSRAYSGQKHDLALTRYQRIIRRLNIFPANKME	303	
Dy	437	nsqlansvmqtllsqkfyepkp-----dlip--pkleaciltggdkis	480	
Qy	304	DVQPIDREWVEEYLLDVLFLNGCRKECASYMANLPVTFR-----YEYL--MA	349	
Dy	481	lqepid-----yllccqlghclawkyntviplqgqeaeaeafyedlddil	527	
Qy	350	ETLFSQILLPQPPF-----KTYLYTLVIMDLCKAL-----PG	382	
Dy	528	esitnmikseledfeldksadfsqstsigknnisatflvmgvcvllnefntsfisfkn	587	
Qy	393	APPAVAVAGVRALEKISDLME-----SRTLILFWSHLSNFOFIWPEWEAFVLDLPK	438	
Dy	598	rfdil--sifmcykkisdlnekgaktkmanktsdillsmk-----	629	
Qy	439	WAPRVVFQEILOREVRVLSYMDKIQSTENATALELLPPKAGPNFMSLEGKKEK----	494	
Dy	630	-----fvssiltalfrllw-----rytsiptsveesgkkgksisllcleglkifsa	679	
Qy	495	-----TEEQQLSABLSRKVRKEKOTARDMIWVI--BETIYPVHGFEV--LTIVVQTLDDTG	546	
Dy	680	vqqfvpkqigflralvtdtkegeeredadvstqtafqirgfrsllnllssgeedfn	739	
Qy	547	SKSFTHLVTVLERYQCVFSKICPDNDKQVLLSQVSTYKNNVQMTAVADRMMGYRLVS	606	
Dy	740	skealllvitvitslkl---lepsspqfvqmlswtskicksredalfckalmnl----	792	
Qy	607	NOAIVRVVFSPENVDFHVS--DQWEILGNALNKNYRISDLRKDKISNTTKNVLVAEKAS	665	
Dy	793	-----lfs-----lhvyskypvillldsqdlnghlqldqgve-----vekn	831	
Qy	666	ANARVEEAESKLSLV---EGEPVLGENPAKMKRLKSTVKTGEAELSRESLEAKAL	722	
Dy	832	hfaivnlrtaaptcvllvsqekveeewdlitlkgvgsqetlseas-----sqatl	885	
Qy	723	LNRALETEVILL--LLFSFGLGVLERLPDPTKVRSDVLKSIGAEDDKPSAMDVDSNG	781	
Dy	886	pnqpvekaimqlgtlllffhelvtalpsgscvdtl--lkd1-----	926	
Qy	782	NPKKSCEVGERQWCLSLTGLYATATROY-----ASEIWPHEKLESEVFSGEDVHPLF	835	
Dy	927	-----ckm-----yttitavryylvcqsgsgipkxmeklvk--lsgshltpic	969	
Qy	836	LQAIS 840		
Dy	970	ysfis 974		

RESULT 14

AM78497	standard; Protein; 710 AA.
ID	AM78497
XX	AM78497;
AC	XX
DT	06-NOV-2001 (first entry)
DE	Human protein SEQ ID NO 1159.
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
XX	nervous system disorder; arthritis; inflammation.
OS	Homo sapiens.
PN	WO200157190-A2.
XX	09-AUG-2001.
PF	05-FEB-2001; 2001WO-US04098.
PR	03-FEB-2000; 2000US-0496914.
PR	27-APR-2000; 2000US-0560875.
PR	20-JUN-2000; 2000US-0598075.
PR	19-JUL-2000; 2000US-0620325.
PR	01-SEP-2000; 2000US-0654936.
PR	15-SEP-2000; 2000US-0663561.
PR	20-OCT-2000; 2000US-0693325.
PR	30-NOV-2000; 2000US-0728422.
XX	(HYSE-) HYSEQ INC.
PI	Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI	Zhao OA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX	WPI; 2001-476283/51.
DR	N-PSDB; AAK51630.
PT	Nucleic acids encoding polypeptides with cytokine-like activities,
PT	useful in diagnosis and gene therapy -
XX	Claim 20; Page 3397-3398; 6221pp; English.
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AAK78323-AAK80302) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC	(AAK80020) are omitted as the relevant pages from the sequence listing
CC	were missing at the time of publication.
SQ	Sequence 710 AA;

Query Match	3.1%;	Score 133.5;	DB 22;	Length 710;
Best Local Similarity	21.0%;	Pred. NO. 0.037;		
Matches 86;	Conservative 74;	Mismatches 163;	Indels 87;	Gaps 16;
Qy	461	KIKQSTENAT-----ALELLPPKAGPNFMSLEGKKEKTEEQQLSALS	505	
Dy	127	kmkqldedasnkaeerarlegelkgldqiaetkarlitqhqhdraqeqdhalmlrelq	186	

Qy	506	RKVEKQTAR-DMTIVJIEET-----IYPVHGFEVTLTITWOTLLDIGSKSFTHLV	554
Dz	187	kllqeertqgdlelrleetrealagrayaaeqmegfelqtqltreveelkse----lq	242
Qy	555	TVLERYGVQSFKLPCPDNDKOVMLLSQSVSTYKKNVNQMTAVADRMGMGYRLVSQAIRVRW	614
Dz	243	airkednqpdpriqlqdgeaarlkshfqalqqcmkrtalaedqlrqgqsveeqrva---	299
Qy	615	FSPENVDPQHVSQDPWEILGN-----ALNKTYNRISDLRKDISNITKNVLVAEKA	664
Dz	300	-alen-----qlsevs-ellgtcyekakqkdlaigklkeri-lqlklenkt---lalaa	347
Qy	665	SANARVELEAAESKLISIVEGEPVLGENPAKMKRCLKTVEKTGAELSLRESLEAKEALLN	724
Dz	348	srsrpldshgeesldv-----nylvdkmekrllygaarksgvtldv-ekicldleimps	402
Qy	725	RAUSETVELLLLLFQPSFLGVLKERLPDPTKVRSVOBCLKSIGAEDDPKPSAMDVDSENGPNK	784
Dz	403	seaadgekatalyyqgelkqlkeef-erykmraqvgvlksknctkd-----gnlg	449
Qy	785	KSCFVGREOWCLSTLCYLTAFTFRQVAS-----EIHPHMEKULESEVFSGE	829
Dz	450	kelaaeaq-----laelkeyisirliscelehqhhdqdaawcke	490

RESULT 15

RESOL 13
ABB67210
ID ABB67210 standard; Protein: 3263 AA.

AA ABB67210;

26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 28422.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
 KW

OS *Drosophila melanogaster*.

PN WO200171042-A2.

PD 27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000: 2000US-191637P

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers FW:

XX
DR WPI: 2001-656860/75.

DR WF1; 2001-030880/
DR N-PSDB; ABL11313;

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure: SEO ID NO 28422: 21pp + Sequence Listing: English

xx CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS57737-ARR27072).

CC The sequence data for this patent did not form part of the printed
CC (ABB51/31-ABB/2012).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence	3263 AA;
SQ	

Query Match 3.1%; Score 133.5; DB 22; Length 3263;
Best Local Similarity 18.7%; Pred. NO. 0.47;
Matches 180; Conservative 142; Mismatches 321; Indels 321;

QY	74	ENEDFVOKLVESVHANFQVALDSGNCNSTRILLRPWTSLCSKVQIPASLIWVF----	ETL	130
Db	2175	esprfveelivqpvemvmdgeal-----lltqvgtgktpkvewyhnaeki	2218	
QY	131	LSAAATTVDEE-KGNPSWQ-----PQADFYVICILSSLPWGSSELAEOVPDEIERVLVGI	184	
Db	2219	tenketisqdgvgvcqlqitvfpeneqgycvatn-----kigksvsktnvki	2268	
QY	185	QAYLSIRKNSSTSGLNFFHNGEFESSLAEKDFVEDLLDRIOQSLASNGWK-LESVP--	240	
Db	2269	qafeyi-pdseitgt-----gseedllrtisideqapklikkipeklep	2313	
QY	241	HLSFEAQLVAGFHEHLPRTKCMQPSPP-----SDHSRAYSKQKHDXALTRYPQRIRLNI	296	
Db	2314	kegeqaki-----evkvvgkpkvkwlrddeqifasee-----	2347	
QY	297	FPANKMED-----VQPIDREVY--EEY-----LLDLVFLVNGC-----RKECA	332	
Db	2348	yqienfedtstvlinhyvpddlgtisfeaynplgvavttalfaveglvgakdyrkpew	2407	
QY	333	SYMANLPVTRVEYLMAETLFSOILLPQPPFKTLYY-----TLVIMDLCKALPGAFPAVV	388	
Db	2408	sqmeemqvalk---diqwyfnvqlrasekyrmvvgqeatelelmkitedscdyyncki	2463	
QY	389	AGAV-----RALPEKISLDMSRTRLILWFSHHLSNFOPTWMBEWAFLVDLPKPAKR	443	
Db	2464	ineigmtmtrakf-----disststiv-----eetkaktvkkksqkk	2501	
QY	444	VVF-----QEILOREVRLSYWD-----KIKOSIE--NATALELLPPKAGPN	483	
Db	2502	tmvkrsgasesqnvktelriptsavetsmnlkvkqpsvkvkseivlvvk----	2557	
QY	484	FMYSLEEGEKTEEQOQSAELSKYKQKTARDMI-----VWIE--ETIYPVHGFE	532	
Db	2558	-----drevadaeersqlieeaeeteekvqhdeedevevqeqetyLsskkie	2610	
QY	533	VLTUI-----VVQTLTD-----IGKSTHVLTVV--L	557	
Db	2611	iktvelirtkisekittiedvgvlshheevqwllesieaesfgigiesalrdiatgll	2670	
QY	558	ERYG-----QVFSKL-CPDNDKQVM-----LLSOVSTYKNNVOMTAV	594	
Db	2671	lyygccheyitmyeqnifislkbpesgalsvqlveregheelisqlisessnedetila	2730	
QY	595	AIDRMNGY-----RLVSNQAIVRWVSPBNVQFHVSQDPWEILGN-----ALNKT	640	
Db	2731	a---tvgfxfpirmqtgyettieivlrkfvredfdis-gdwkicgkeeeilpqvlainee	2786	
QY	641	YNRISD--LRKOI-----SNITKNVLVAEKASANARVELEAAESKLSLVEGEPVLGNPA	693	
Db	2787	vpldfeigrkesprkpenkiltienivrlkhalnvs-haktaeeskelpskipksvkaqr	2845	
QY	694	KMKRLKSTVEKTGEAFLSD-----RESLEAKEALLNRA	726	
Db	2846	kmkesrlsvaeapaeatedlkplkavseqvsdillfshetteeqhqaletieklkpts	2905	
QY	727	LSFTEVLLLLPQSFGLVKERLPDPTKVRVSQQDLKSTIGAEDDKPSA---MDVDSNGNP	783	
Db	2906	atedtvqgklisgeel-iaaevlpsaetvgrdvtvrrpg-etisprltpmnslicetcp	2963	
QY	784	KKSCEVGEREQNCLSTLGLYTAFTROYASEIWPWHKEULESEVFGEDVHPLQOAISSAL	843	
Db	2964	eds--lqenmq-----aakermet-----pmsvteskavagqge-----levlenvd	3003	

Oy 844 QFPL 847
||
Db 3004 hmpl 3007

Search completed: August 20, 2002, 15:25:52
Job time: 5641 sec